JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 20, No. 11 NOVEMBER, 1949

In This Issue

		Page
Analysis of the Mechanics of Metal Cutting	D. C. Drucker	1013
Antennas and Open-Wire Lines. III. Image-Line Measurements	Patrick Conley	1022
Convection Currents in Porous Media. II. Observation of Conditions at Onset of Convection	H. L. Morrison, F. T. Rogers, Jr., and C. W. Horton	1027
Direct Method of Determining Preferred Orientation of a Flat Reflection Sample Using a Geiger Counter X-Ray Spectrometer	L. G. Schulz	1030
Determination of Preferred Orientation in Flat Transmission Samples Using a Geiger Counter X-Ray Spectrometer	L. G. Schulz	1033
Propagation of Transverse Waves in Viscoelastic Media	F. T. Adler, W. M. Sawyer, and John D. Ferry	1036
Study of the Mechanism of Sintering of Metallic Particles	J. H. Dedrick and A. Gerds	1042
Harmonic and Subharmonic Response for the Duffing Equation $\ddot{x}+\alpha x+\beta x^3=F\cos\omega t(\alpha>0)$	Morris E. Levenson	1045
New Interpretation of the *n*-Power Law in Plastic Deformation	Herbert I. Fusfeld	1052
Technique for the Preparation of Grease or Solid Samples Dispersed in Grease-Like Media for Examination with the Electron Microscope	A. Y. Mottlau	1055
Increasing Space-Charge Waves	J. R. Pierce	1060
Effect of Cold Work in Metals on Powder Pattern Intensities	B. L. Averbach and B. E. Warren	1066
Structure of Porous Materials from Gas Penetration Rates	F. A. Schwertz	1070
Application of a Variational Principle to Biconical Antennas	C. T. Tai	1076
Erosion of Electrical Contacts on Make	L. H. Germer and F. E. Haworth	1085
Appendix A. Effect of Steady Value Current upon Erosion	D. P. Ling	1108
Scattering Cross Section of Spheres for Electromagnetic Waves	L. Brillouin	1110
Measurement of the Complex Conductivity of an Ionized Gas at Microwave Frequencies	Fred P. Adler	1125
Letters to the Editor:		
Thermionic Emitting Properties of Two Refractories	R. E. Haddad, D. L. Goldwater, and F. H. Moran	1130
Papers by Hadrian D. Stahl	C. H. Metson	1130
Index to Advertisers		xxvi

BOARD OF EDITORS

Elmer Hutchisson, *Editor*

Associate Editors

R. Bowling Barnes
Otto Beeck
W. James Lyons
Philip Morrison
Paul E. Sabine
Raymond J. Seeger
Frederick Seitz
Charles S. Smith
J. R. Townsend

Publication of the American Institute of Physics

OFFICERS

George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*
John T. Tate, *Adviser on Publications*

GOVERNING BOARD

Luis W. Alvarez
J. W. Beams
Wallace R. Brode
E. U. Condon
W. F. Fair, Jr.
Harvey Fletcher
George R. Harrison
W. V. Houston
Paul Kirkpatrick
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
J. R. Oppenheimer
George B. Pegram
Harold K. Schilling
John T. Tate
Wallace Waterfall
M. W. White
A. G. Worthing

Consultant on Printing

Melvin Loos

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Business Manager*
Ruth F. Bryans, *Publication Manager*
Betty Hofstadter, *Office Manager*
Eileen B. Neuberger, *Circulation Manager*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. The authors' institutions are requested to pay a **publication charge** of $4 per page which, if honored, entitles them to 100 free reprints. Instructions will be sent with galley proofs.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$7.00	$ 8.00
To all others	9.00	10.00

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 19, 1948—$140.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1949, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 20, Number 11 November, 1949

An Analysis of the Mechanics of Metal Cutting*

D. C. Drucker**
Armour Research Foundation, Chicago, Illinois
(Received January 10, 1949)

An attempt is made to go beyond an idealized description of the process of metal removal. The details of two-dimensional orthogonal cutting are investigated, and several concepts and hypotheses are introduced which apply equally well to the most general cutting operation. Of these, the most significant is the great importance of the small time available for the plastic deformation to occur. Ekstein's paradox is explained, and the influences of speed of cutting, depth of cut, and rake angle are correlated by the consideration of dynamic plasticity. In addition, the effect of non-homogeneity of the material is analyzed and shown to account for the gradual change from the discontinuous to the continuous type of chip as the cutting speed is increased, and for the marked increase in the specific cutting force as the depth of cut is decreased. A partial theoretical analysis of the state of stress, in the chip and the workpiece, and energy considerations are also found helpful in the description of what probably happens when metal is cut.

INTRODUCTION

IN this discussion of the process of metal removal, the initial reference picture to which all quantities will be referred is that of an ideal material, perfectly homogeneous with the same strength and flow properties in all directions. The idealizations will be studied and some will be removed in the analysis subsequent to the following descriptive outline.

In the assumed two-dimensional cutting operation, as shown in Fig. 1(a), metal is removed in the form of a continuous chip without a built-up edge. For convenience, the tool of rake angle α is considered as stationary and the workpiece is assumed to move past it with a speed V. As soon as the metal reaches the shear plane OA, and not until then, it is sheared and moves on without further distortion. Each and every plane of metal is displaced in this manner as a deck of infinitesimally thick cards would be (Fig. 1(b)). This admittedly simplified but very valuable type of description was advanced by Piispanen,[1] and, independently, by Ernst and Merchant.[2,3] The shearing stress on the shear plane, τ_y, may be considered either as a definite number or as a function of the normal stress σ. A free body sketch of the chip is shown as Fig. 2. The force exerted on the chip by the tool is R. It is inclined with respect to the normal to the tool chip interface by an angle β, a "friction" angle which is assumed to remain constant under various conditions of cutting.[4]

The geometry of the cutting action in this idealized theory is determined by postulating that the shearing action will take place on the plane which will require the least magnitude of H, the effective cutting component of R. If τ_y is independent of the normal stress on the plane (Bridgman has shown that the variation is not very large) then the angle ϕ is the angle at which the shearing stress $F_s/(bt/\sin\phi)$ is maximum. In general, however, it is found[5] that ϕ must satisfy

$$2\phi+\beta-\alpha=C \qquad (1)$$

* The results presented were obtained during an investigation sponsored by the Shell Oil Company.

** Now Associate Professor of Engineering, Brown University, Providence, Rhode Island.

[1] V. Piispanen, "Lastunmuodostumisen Teoriaa" (Theory of chip formation), Teknillinen Aikakauslehti 27, No. 9, 315–322 (1937).

[2] H. Ernst and M. E. Merchant, "Chip formation, friction, and finish," The Cincinnati Milling Machine Company (1940). (Also, "Surface treatment of metals," Symposium, Trans. A.S.M.E., p. 299–378 (1941).)

[3] H. Ernst, "Physics of metal cutting," The Cincinnati Milling Machine Company, 1938. (Also, "Machining of metals," Trans. A.S.M.E., p. 1–34 (1938)).

[4] This assumption, implicit in Piispanen's work, was dropped by Ernst and Merchant, but no theory of the variation of β was given.

[5] M. E. Merchant, "Mechanics of the metal cutting process. I. Orthogonal cutting and a type 2 chip," J. App. Phys. 16, 267–275 (1945). "II. Plasticity conditions in orthogonal cutting," J. App. Phys. 16, 318–324 (1945); "Basic mechanics of the metal cutting process," J. App. Mech., Trans. A.S.M.E. 66, A-168–175 (1944).

where $C = \text{arc ctn}k$; k is the slope of the required shear stress *versus* normal stress curve, and, therefore, $C = \pi/2$ when τ_y is independent of σ.

The force R is simply

$$R = \frac{F_s}{\cos(\phi+\beta-\alpha)} = \frac{\tau_y bt}{\sin\phi\cos(\phi+\beta-\alpha)}. \quad (2)$$

As can be seen from the preceding development, this very simplified description of two-dimensional cutting requires only C, α, and β for the determination of the geometry, that is, ϕ and τ_y in addition for the value of R/bt, or the specific cutting force $H/bt = (R/bt)\cos(\beta-\alpha)$. There is no "scale effect," no influence of speed, no detailed consideration except for the effect of the normal stress on τ_y. Many modifications are required for a more complete theory which will account for at least some of the observed effects.

A few of the interesting and important facts of cutting which should be explained by a reasonably complete theory are the following:

1. *Ekstein's Paradox of Cutting.*[6]—The stresses are extremely high, of the order of the ultimate static strength, in a large region of the chip and workpiece, and yet the zone of shearing is extremely narrow (Fig. 6).

2. *Changes in Continuity of the Chip.*—As the cutting speed V increases, or as the rake angle α increases, or as the depth of cut t decreases, there is a change from a discontinuous to a continuous type of chip. It is approximately true that all materials behave more plastically as the speed of cutting is increased.

3. *Specific Cutting Force.*—At small depths of cut, the required force per unit of area increases appreciably as the depth of cut decreases (Fig. 12).

4. *Depth of Cold Work in Workpiece.*—The region near the new surface of the workpiece is not undisturbed; on the contrary, there is noticeable plastic flow in the workpiece to a depth approximately equal to the depth of cut.

SYMBOLS AND FORMULAS

b = width of workpiece (cut) which is very large compared to t.
C = arc ctn$k$.
F = frictional component of $R = R\sin\beta$.

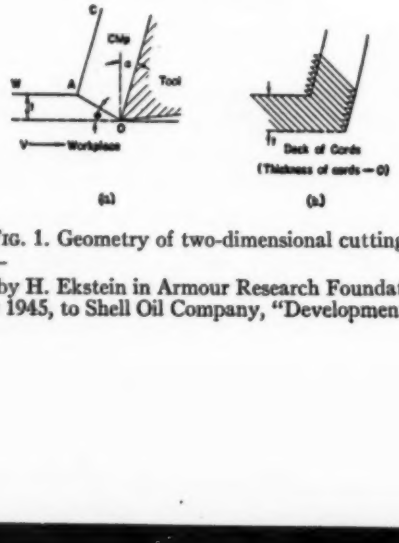


FIG. 1. Geometry of two-dimensional cutting.

[6] Stated by H. Ekstein in Armour Research Foundation Report No. 9, July 1945, to Shell Oil Company, "Development of cutting fluids."

F_s = shearing force on the shear plane $= R\cos(\phi+\beta-\alpha)$.
G = shear modulus.
H = power component of $R = R\cos(\beta-\alpha)$.
k = slope of shear flow strength *versus* normal stress curve $\tau_y = \tau_0 + k\sigma$.
N = normal component of $R = R\cos\beta$.
N_s = normal force on shear plane $= R\sin(\phi+\beta-\alpha)$.
P = force in the vicinity of the tool tip.
R = force exerted on the chip by the tool.
r, θ = polar coordinates.
s = distance traveled by workpiece relative to tool.
t = depth of cut.
t' = thickness of shear zone.
V = cutting speed.
V' = component of V perpendicular to shear plane $= V\sin\phi$.
x, y = Cartesian coordinates.
α = rake angle.
β = angle between R and the normal to the chip, $\tan\beta = F/N$.
γ = shear strain $= \cos\alpha/\sin\phi\cos(\phi-\alpha)$.
γ_L = a limiting value of shear strain.
ϕ = shear angle.
ρ = mass density.
σ = normal stress.
τ = shearing stress.
τ_y = yield stress in shear on the shear plane $= F_s/(bt/\sin\phi)$.

EKSTEIN'S PARADOX OF CUTTING

Hardness measurements,[7] photo-micrographs,[2,3,5] and visual observation of cutting show that the enormous plastic deformation of the chip takes place in a very narrow zone of shear, $OABO$, as shown schematically in Fig. 3. Static reasoning would predict a very large increase in the shearing stress (see Fig. 4) in going from the workpiece, which is practically undeformed, through the zone of thickness t', into the chip in which the permanent shear strain exceeds unity and is often four or more. This increase, on the basis of a static stress-strain curve, would be approximately 20,000 p.s.i. for mild steel.

As shown in the following paragraphs, Newton's laws of motion would be contradicted if the increase were even a small fraction of 20,000 p.s.i. This is the paradox of cutting as stated by H. Ekstein:[6] large increase in strain with very small if any change in stress. After consideration of the paradox, it will be restated in an equivalent form more suitable from the present physical point of view.

When the tool is stationary and the work moves by it, the motion and shearing action is (on the average) steady in time, and Newton's second law (force equals

[7] N. Zlatin and M. E. Merchant, "The distribution of hardness in chips and machined surfaces," A.S.M.E., Advance Paper for Semi-Annual Meeting, June 1946, published as a special pamphlet, "Cutter life, hardness distribution, and honing," included with Trans. A.S.M.E. 69 (1947).

time rate of change of momentum) becomes simply, resultant force equals mass per unit of time, multiplied by the change in velocity. From the free body sketch (Fig. 3(b)) with b the dimension perpendicular to the paper

$$\Delta F_s + P = (\rho V b t)\Delta V \tag{3}$$

where ρ is the mass density, ΔF_s is the increase in shearing force across the shear zone, and ΔV is the vector change in velocity whose magnitude is

$$\Delta V = V\frac{\cos\alpha}{\cos(\phi-\alpha)}. \tag{4}$$

As ΔV will not exceed $2V$ and the force P exerted in the vicinity of the tool tip is probably compressive as shown,

$$\Delta F_s < 2\rho b t V^2. \tag{5}$$

The increase in the average shearing stress

$$\Delta\tau = \frac{\Delta F_s}{bt/\sin\phi} < 2\rho V^2 \sin\phi < 2\rho V^2 \tag{6}$$

at the high cutting speed of 1000 feet per minute is therefore less than

$$2\left(\frac{490}{32.2}\right)\left(\frac{1000}{60}\right)^2 = 8500 \text{ pounds per square foot}$$

or about 60 p.s.i., an entirely insignificant figure.

Whether P is tensile or compressive, its effect on $\Delta\tau$ is small because the shear zone is so extremely narrow compared with its length.

The same result can be obtained by the use of the differential equations of motion of the theory of elasticity. It is clear that curvature of the shear surface will not change the answer appreciably.

Therefore, in traversing the zone of shear the material is subjected to strains of a few hundred percent, but the shear stress at the beginning, the end, and throughout the zone is approximately the same.

A plausible explanation of this paradox is that the stress-strain diagram is not the customary static one, but is instead similar to the "ideal" one postulated in elementary plasticity theory (Fig. 5).

The true physical picture of the plastic behavior requires consideration of experimental values for stress magnitude as well as gradient. Experimenters and production men agree that the specific cutting force, H/bt, is usually between 250,000 and 600,000 p.s.i. for steel, as compared with the nominal tensile ultimate of 60,000 to 150,000 p.s.i. *The average shearing stress on the plane of shear OA (Fig. 6), is much larger than the nominal tensile strength.* The average normal stress on the shear plane at usual cutting speeds and rake angles is approximately twice the shear stress.

Consideration of these numerical values does not change the essential nature of the paradox, but does lead to a restatement and re-evaluation. The large values of normal and shearing stress on OA mean that high shearing stresses must exist throughout a large region, for example, as shaded on Fig. 6.

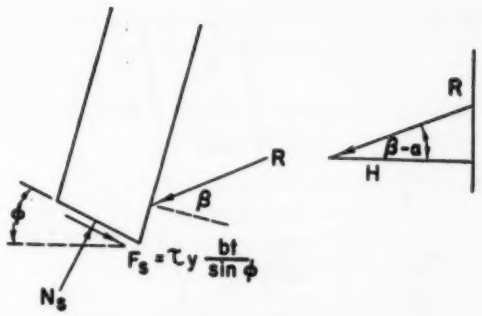


FIG. 2. Chip as free body.

The paradox from this point of view is that a region under OA (in the workpiece) has shearing stresses averaging more than three times the static yield point in shear, and yet plastic flow does not occur. Restated in this form, it becomes clear that the concepts of quasi-static plasticity must be replaced by a dynamic theory.

THE STATE OF STRESS IN CHIP AND WORKPIECE

An exact solution for the stresses in the chip and the workpiece would require a study of the chip, the shear zone, the elastic region of the workpiece directly below the shear zone, and the plastic region behind the tool point. The geometry of the parts involved is fairly simple, but it is improbable that a solution could be obtained in closed form even if temperature effects are neglected. However, some general conclusions can be drawn that are of interest and which can serve as a check on the various hypotheses employed in the study of metal cutting.

Considering the chip, the shear zone which may be curved, and the workpiece (Fig. 7), the continuity conditions require that the normal and shear stresses and the normal and tangential displacements be the same at OA in the chip and OA in the shear zone, and also at $O'B$ in the shear zone and $O'B$ in the workpiece. As the shear zone is very narrow, the stresses on OA and $O'B$ must be very closely the same in order to avoid a contradiction with Newton's laws. The normal displacements at the two long bounding surfaces must also be almost the same. Only the tangential displacements can differ appreciably. The difference between them will be the shearing strain multiplied by t', the zone thickness. Therefore, as the zone gets narrower, the stresses and displacements for OA and $O'B$ differ less and less, and in the limit as t' goes to zero they become the same except for a very small difference between τ_C and τ_W of order ρV^2, as discussed in the previous section on the cutting paradox.

Therefore, in the ideal or limiting case the effect of the shear zone on the stresses in the chip and workpiece may be neglected and in general it must be small. Stresses

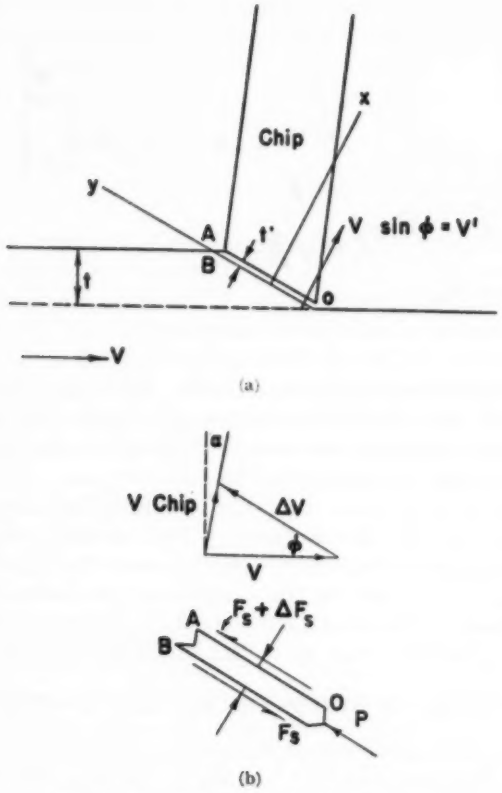


FIG. 3. Change in stress across shear zone.

and displacements are continuous across the zone. Both the chip and the workpiece near the shear plane are in the elastic state, and the stress problem is reduced to a problem in elasticity complicated, however, by the very uncertain effect of temperature on strain and material properties. This means that for the ideal case where temperature effects and the plastic flow in the workpiece near the tool tip can be neglected,[8] the state of stress and the loading of the chip by the tool can be determined experimentally. If the shear and normal stresses on the (tool-chip) boundary of an elastic model are adjusted individually point by point to make the shear line OA, a line of (almost) constant shear stress, and at the same time a trajectory of shear stress, then the stresses throughout the chip and workpiece and also the boundary loading will be those existing in the ideal case when metal is cut with a shear angle and a rake angle equal to those of the model. A photo-elastic experiment employing either gelatin or a very sensitive plastic, such as Marblette, is indicated. If the shear line is straight, it will be an isoclinic as well as a stress trajectory.

The boundary loading cannot be obtained by pressing a simulated tool against the chip boundary because the dynamic normal and shearing stresses are entirely different from those that would be obtained statically at the interface. It cannot be obtained by an actual cutting operation because the highly strained chip cannot be analyzed photo-elastically.[9]

Temperature stresses in usual lathe or milling operations will be quite large,[8] but the total stress is so high that except for local effects the general pattern of stress and boundary loading should be given reasonably well. The influence of the plastic flow that takes place in the workpiece as it passes the tool point is more troublesome. An analysis cannot be given until the ideal case has been solved.

Some conclusions can be drawn without performing an experiment. Shearing occurs all along the line OA, and therefore, OA will be a line of maximum shear stress[10] and also a trajectory of shear stress; that is, the two normal stresses σ_n and σ_t are equal and the derivative of the shearing stress τ_{nt} with respect to n or t is zero (Fig. 8). If OA is a straight line (Fig. 3), then the equations of equilibrium

$$\frac{\partial \sigma_x}{\partial x}+\frac{\partial \tau_{xy}}{\partial y}=0, \qquad \frac{\partial \tau_{xy}}{\partial x}+\frac{\partial \sigma_y}{\partial y}=0, \tag{7}$$

reduce to

$$\frac{\partial \sigma_x}{\partial x}=0, \quad \frac{\partial \sigma_y}{\partial y}=0 \tag{8}$$

so that $\sigma_x=\sigma_y=$ a constant along the shear plane except for discontinuities possible at the end points O and A, and σ_x is a maximum at the shear plane.

Therefore, both F_s and N_s (Fig. 2) are uniformly distributed along OA. For equilibrium, R, the force exerted by the chip on the tool must pass through the midpoint of the shear plane (Fig. 9(b)).

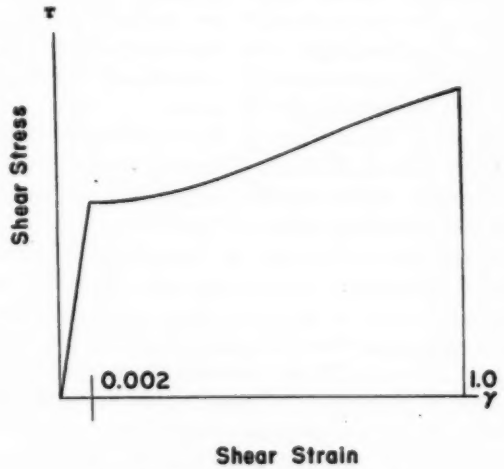


FIG. 4. Static stress-strain diagram.

[8] Temperature changes are small at low speeds and small depths of cut.

[9] E. G. Coker and L. N. G. Filon, *A Treatise on Photo-Elasticity*, (Cambridge University Press, London, 1931), p. 336–351.

[10] Not a maximum for all possible ϕ, but a maximum in the given geometrical configuration.

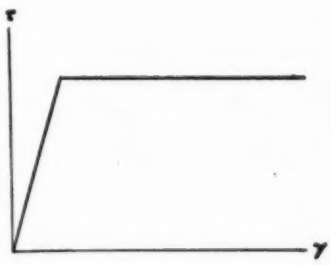


FIG. 5. "Ideal" plasticity.

The effects of curvature of the shear surface can be determined qualitatively from the equations of equilibrium in polar coordinates (Fig. 8),

$$\frac{\partial \sigma_r}{\partial r}+\frac{1}{r}\frac{\partial \tau_{r\theta}}{\partial \theta}+\frac{\sigma_r-\sigma_\theta}{\partial r}=0,$$
$$\frac{\partial \tau_{r\theta}}{\partial r}+\frac{1}{r}\frac{\partial \sigma_\theta}{\partial \theta}+\frac{2}{r}\tau_{r\theta}=0, \tag{9}$$

which reduces to

$$\frac{\partial \sigma_r}{\partial r}=0, \quad \frac{\partial \sigma_\theta}{\partial \theta}=-2\tau_{r\theta}$$

or

$$\sigma_\theta=(\sigma_\theta)_a-2\theta\tau_{r\theta} \tag{10}$$

which means that convex curvature (Fig. 9(a)) corresponds to an algebraic decrease in normal stress or increase in compressive stress going from A to O, and concave curvature (Fig. 9(c)) corresponds to a decrease in the compressive stress.

Since the curvature of the shear surface is small along practically all of its length, the force R must pass close to the midpoint of the surface (Fig. 9).

HYPOTHESES TO EXPLAIN DETAILS OF SHEARING ACTION

Relatively little is known about flow and fracture of material for simple states of stress and quasi-static loading. The understanding of plastic behavior under the extreme conditions of metal cutting is far more severely limited. The following speculations employ known facts but may in the future be found physically invalid in many respects. Their justification is that they supply a useful and consistent basis for qualitative explanation of the observed phenomena.

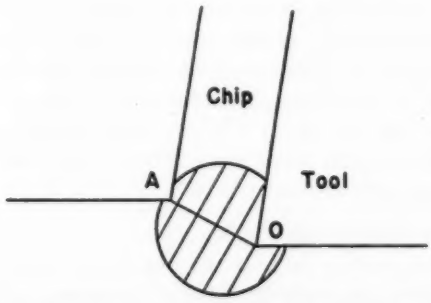


FIG. 6. Zone of high shearing stress.

A. Explanation of the Paradox of Cutting—A Time Effect

So far, the time variable has been completely omitted from consideration. A rough calculation of the rate of shear strain shows its importance. The average rate of shear strain is equal to the shear strain γ divided by the time required to traverse the shear zone, t'/V' (Fig. 3). Taking a cutting speed of 200 feet per minute and a depth of cut of 0.010 inch, observing from photomicrographs that the thickness of the shear zone is at most one-twentieth of the depth of cut and the shear angle generally exceeds 15 degrees, the time of traverse is less than

$$\frac{(0.010/20)}{(200)(12/60)(0.25)}=0.00005 \text{ second},$$

which is extremely short. For a shear strain of two, the rate of strain would exceed 40,000 per second. Only for very low cutting speeds and deep cuts does the strain rate go down to the range of the very high speed (impact) tension tests made by Manjoine[11] and by Fehr.[12]

These enormous rates of strain are the reason why the static stress-strain curve is not correct in shape or magnitude of predicted stress. A possible method of explanation is that under static conditions, as the load is increased, dislocations travel to planes of weakness and at a certain stress called the yield stress, slip will occur on enough planes to be obvious. After slip, the weak planes are strong and an increase in stress is required for slip to occur on another set of planes on which weaknesses accumulate. These planes slip and in turn become strong, requiring a further increase in stress to get slip on the next stronger set, and so on. Some such process of work hardening accounts for the positive slope of the stress-strain curve in the plastic region (Fig. 4).

At high rates of strain there is no opportunity for the weaknesses to migrate appreciably, and so to accumulate and to lower markedly the yield strength of a large number of planes. There is also no opportunity for

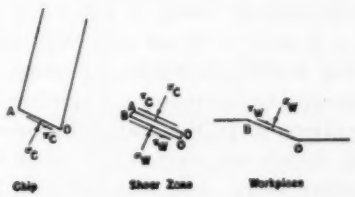


FIG. 7. Boundary stresses on chip, shear zone, and workpiece.

[11] M. J. Manjoine, "Influence of rate of strain and temperature on yield stress of mild steel," J. App. Mech., Trans. A.S.M.E. 66, A-211–218 (1944).

[12] Fehr, Parker, and DeMicheal, "Measurement of dynamic stress and strain in tensile test specimens," J. App. Mech., Trans. A.S.M.E. 66, A-65–71 (1944).

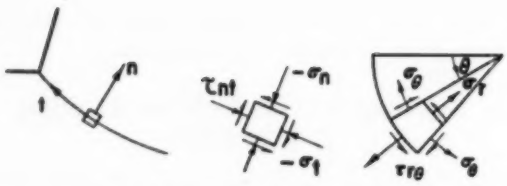


FIG. 8. Stresses at shear surface.

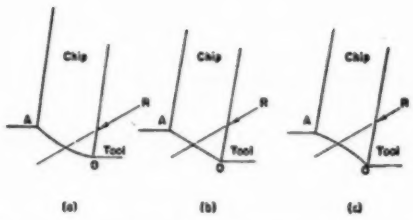


FIG. 9. R acts along a line which passes close to midpoint of shear surface.

dislocations to be regenerated. The dynamic stress required for slip is therefore far above the static, and the difference between planes is much less pronounced. Figure 10 shows the probable stress-strain curves at various rates of loading. The tests by Manjoine and by Fehr, previously referred to, lend substantiation to this hypothesis, although they were tension tests and were made at strain rates which are very high but still small compared with those encountered in machining operations. The best check is given by Manjoine's observation that at room temperature, the yield point and the ultimate coincide at high strain rates. An increase in temperature makes for an additional difference between yield stress and ultimate stress which would probably diminish and disappear if the strain rate were raised sufficiently, as increase in strain rate and decrease in temperature are somewhat equivalent.

It seems probable, therefore, that the ideal plasticity curve actually does exist, and that the value of the dynamic yield stress is above the static ultimate. The paradox of cutting is resolved in that the plane or surface of shear is one of high shear stress, and slip does not occur on surfaces more than a small distance away because the stress on these surfaces is not quite high enough for the time available. This is an oversimplification, because the potential shear surface is in a region of high stress as soon as it enters the region represented by the shaded portion of Fig. 6, so that the over-all time at the various stress levels should be considered. The conclusion is, of course, qualitatively the same.

The higher ductility compared with the familiar static tests is partly due to the high rate of strain (energy released at the slip planes does not have time to dissipate to the surrounding metal), but probably is mainly the result of the high all-around compression, Fig. 11, in the zone of shear. Hardening still occurs despite the absence of a "work-hardening" range in the static sense.

Perhaps it is well to repeat and emphasize here that metal cutting is not cutting in the usual sense, but is instead a removal by brute force. There are large stresses and deformations in all the material removed, while in true cutting, bonds are severed in a small region of the material taken away, and relatively little energy is required.

B. Change from Continuous to Discontinuous Chip

The simplified description of the physical properties of metals permits the explanation of their actual behavior during cutting, and gives the deviations from the idealized picture, Fig. 1. As the rake and shear angles determine the shear strain γ, the idealization can be valid only if γ does not exceed some limiting value of ductility, γ_L. This limit will depend upon many factors; among them are the hydrostatic pressure or average of the principal stresses, the temperature, and the time available for fracture.

The variation of the ductility with the normal stress on the shear plane is not known for appreciable rates of strain. It is possible that there be little or no variation, but much more likely that, as in Bridgman's tests under static conditions, the higher the normal stress the larger the plastic flow that can take place before fracture occurs. The normal stress just mentioned is really a uniform all-around compression (Fig. 11) as $\sigma_n = \sigma_t$ (Fig. 8), and Poisson's ratio is closely $\frac{1}{2}$.

The effect of temperature is more complicated. For the initial range of a few hundred degrees rise above room temperature, the pattern is definite for each metal but the ductility may increase, decrease, and then increase again. In the higher range this irregularity disappears and the increase is large.

Some experimental observations on tensile specimens of metal support the importance of a time for fracture which is often observed in plastics. If the specimen is stretched a very large amount at a high rate of strain and then held in the strained position, tensile failure will occur after an appreciable time has elapsed. In the case of metal cutting, "appreciable" is to be interpreted as of the order of the 0.00005 second computed previously, or larger. The larger the strain the less the time required, but a lower limit may very well exist. Conversely, the less the time in the shear zone, the higher the strain which can be imposed without fracture.

The idea of a limiting shearing strain also must involve a probability concept. In cutting, all regions of the material being removed are subject to large distortion at extreme rates of strain, and the non-homogeneity of the material is significant. The limiting strain is not the same for all the shear planes; some planes are considerably weaker, some considerably less able to slip large distances without fracture.

1. *The Effect of the Cutting Speed.*—It is not actually possible to isolate each of the important variables and study them separately because of the strong influence of each on the others. However, for clarity, it will now be

supposed that the forces and geometry are constant while the cutting speed alone is changed.

At some very high speed, the time available for fracture is so small that plastic deformation without fracture may possibly occur on all planes sheared for any rake angle. At a high, commercially available speed, only partial fracture will occur on any plane and then so infrequently as to be unimportant except possibly to enable chip breakers to be effective. The chip will come off continuously as the speed is decreased, but at some value of V the partial fractures will be sufficiently deep and closely spaced to be noticeable. γ_L could be defined somewhat indefinitely by the appearance of this type of chip. At lower speeds still, the fractures would become practically complete and a segmental chip would result. The transition from the continuous chip to the discontinuous is gradual. On the other hand, if the metal were completely homogeneous with a sharp value of γ_L there would be an abrupt transition—a continuous chip down to some lower limit of velocity and then, suddenly, complete disintegration.

2. *The Effect of Rake Angle.*—As the rake angle is decreased, the shearing strain required for continuous chip formation increases. Therefore, if the normal stress and the temperature on the shear plane did not change, a higher cutting speed would be and generally is required for the smaller rake angles. However, the normal stress increases and it is possible that the resulting effect at large negative rake can more than make up for the large strain required.

3. *The Effect of Depth of Cut.*—The time available for fracture is approximately proportional to the depth of cut and inversely proportional to the cutting speed. Therefore, lower cutting speeds are required at smaller depths of cut to obtain the same degree of chip continuity.

C. The Effect of Depth of Cut on Specific Cutting Force

Another result of the consideration of the details of cutting, including the non-homogeneity of the material, is the prediction of the variation of the specific cutting force with depth of cut.

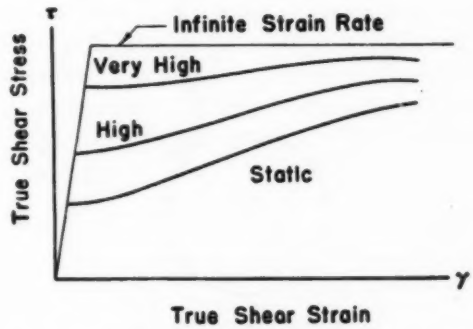


FIG. 10. Probable stress-strain curves as a function of strain rate. The total γ to fracture (ductility) depends on the normal stress, and increases with increasing all around compression.

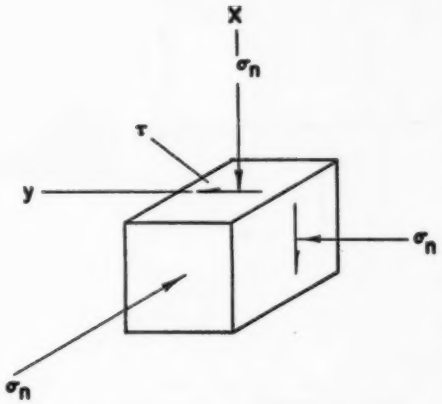


FIG. 11. Hydrostatic pressure plus shear.

Figure 12(a) shows a material containing a large number of points of weakness scattered at random. Only a few of these points are drawn. The rake angle α and the shear plane making an angle ϕ with the direction of cutting are also indicated. For simplicity of description it will be assumed that the shear angle is the same for all depths of cut (Fig. 12(b) and 12(c)).

Shearing occurs on a very large number of planes, but not necessarily or actually on all. As *the total amount of slip required for the geometry of cutting is independent of the depth of cut*, the larger the depth of cut is, the more the planes of weakness available, and as a consequence, the less the average of the shearing stresses on the shear planes. After a sufficient depth of cut is reached, there are an ample number of the weaker planes for the geometry, and further increase in depth has no influence because the strength per unit area is practically constant.

The fewer the weaker planes the higher the average shear stress and the larger the resultant force per unit of area R/bt or the specific cutting force H/bt. This phenomenon is similar to the size effect so often observed in fatigue tests where the smaller a specimen with a stress concentration, the higher the fatigue limit, or the longer the life at a given nominal stress.

The same effect would also be produced by a tool with a rounded instead of a sharp edge, used at different depths of cut. However, the curve of Fig. 12(d) is observed for extremely sharp tools in the range where such rounding is negligible.

D. Depth of Plastic Flow in the Workpiece

The state of stress throughout the chip and the workpiece is determined by the tool loading on the chip and, also, by the temperature, which will again be ignored. This loading is unknown, but in the previous discussion of stress it was shown that the resultant force R passes close to the midpoint of the shear plane and, therefore, intersects the tool chip interface at a distance from the tool point which is approximately equal to the depth of cut.

An idea of the magnitude of the stresses produced can

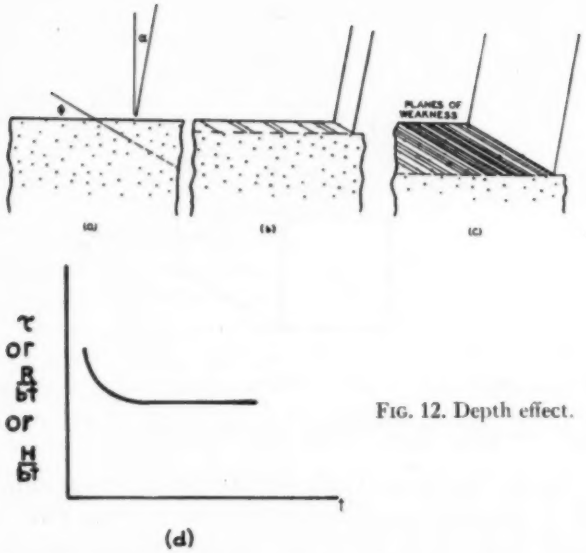


FIG. 12. Depth effect.

be obtained from the radial stress distribution in an elastic solid subject to a concentrated force R at the boundary (Fig. 13). The contours of constant stress are circles passing through the point of application of the load with centers on the line of action of the load. The magnitude of the stress at each point of such a circle of diameter d is approximately $2R/1.5\pi$ db or $(R/bl)(2l/1.5\pi d)$, which for mild steel is close to $150{,}000l/d$, or 75,000 p.s.i. for $d=2l$. Therefore, since the actual force is distributed, it is reasonable to expect that both tensile and compressive stresses somewhat exceeding the static yield stress will exist in the workpiece at a depth equal to the depth of cut. Until a good experimental or theoretical solution is obtained for the boundary loading, there is no point in trying to be more precise about these stresses.

The time available for yielding to take place in the zone subject to these tensile and compressive stresses is much longer than for the action on the shear plane of the chip, and yielding can take place at the lower stresses. However, it will not progress far because the time is still short, and the geometry of cutting does not require large deformations.

ENERGY CONSIDERATIONS

The work input, H times the distance s traveled by the workpiece relative to the tool, is dissipated in shearing the metal removed, in friction between the tool and the chip, and in deformation of the workpiece. Although the idea of an energy balance is clear, there are some subtle and possibly significant points involved. As will be seen in the following derivation, a pseudoenergy balance can be obtained from equilibrium and geometrical considerations only.

By definition (Fig. 2), $H=R\cos(\beta-\alpha)$, and therefore no matter what s denotes, $Hs=sR\cos(\beta-\alpha)$. Also, an expansion will verify that for any angle ϕ

$$\cos(\beta-\alpha)\equiv\frac{\sin\beta\sin\phi}{\cos(\phi-\alpha)}+\frac{\cos(\phi+\beta-\alpha)\cos\alpha}{\cos(\phi-\alpha)}$$

so that

$$Hs=\frac{sR\sin\beta\sin\phi}{\cos(\phi-\alpha)}+\frac{sR\cos(\phi+\beta-\alpha)}{\cos(\phi-\alpha)}$$

for any s and any ϕ. If, now, ϕ is identified with the shear angle, and the geometry of cutting is assumed, see Fig. 14 and list of symbols and formulas

$$Hs=\frac{Fs\sin\phi}{\cos(\phi-\alpha)}+\frac{F_s\gamma(sbt)}{bt/\sin\phi}. \tag{11}$$

If, furthermore, s is chosen to be the distance traveled by the workpiece relative to the tool, Eq. (11) can be interpreted term by term as work input = (frictional force) (relative motion between chip and tool) + (shearing stress on shear plane) (shearing strain in chip) (volume of metal removed). However, this equation is true solely because of the conditions of equilibrium and geometry, and in no way involves actual work or energy considerations. The "energy balances" reported[5,13] are simply checks on numerical computations.

Assuming a simple rubbing action between chip and tool, the true work equation reads work input = (frictional force) (relative motion between chip and tool) + (work per unit volume required to shear the metal removed) + (work required to deform the workpiece).

Therefore, by comparison with Eq. (11) the experimental observation that the last term is small compared with the first two means that the work per unit volume to shear the metal removed is almost equal to the product of the shearing stress on the shear plane and the maximum shearing strain produced. As the general expression for the work per unit volume is $\int_\gamma \tau d\gamma$, the requirement that it be almost $\tau\gamma_{\text{maximum}}$ provides another proof that, under continuous cutting conditions, metal behaves closely in the ideally plastic manner of Fig. 5.

It is interesting to note that if the ideal curve is followed exactly, the necessary work of deformation per unit volume is $\tau_y^2/2G$ lower than $\tau\gamma_{\text{maximum}}$. Therefore,

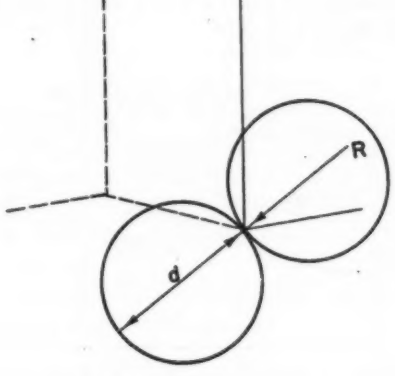


FIG. 13. Stress contours.

[13] M. E. Merchant and N. Zlatin, "New methods of analysis of machining processes," Proc. S.E.S.A. 3, No. 2, 4–27 (1946).

if the deformation of the workpiece involves only this small an average work per unit volume in depth t, the ideal curve must actually hold; for the cutting of mild steel,

$$\frac{\tau_y^2}{2G}=\frac{(80{,}000)^2}{(2)(12{,}000{,}000)}=270 \text{ inch-pounds per cubic inch}$$

which is ample on a static basis for the average plastic strain actually observed.

The foregoing discussion shows that the stress-strain curve must be very near to the ideal one. Then quite closely the ratios, shear work: friction work: total work are:

$$[\cos(\phi+\beta-\alpha)\cos\alpha]:\sin\beta\sin\phi:[\cos(\beta-\alpha)\cos(\phi-\alpha)]. \quad (12)$$

As an example, if $\beta=45°$, $\alpha=10°$, $\phi=20°$, the ratios are 0.7:0.3:1.0 and the shear deformation at the shear plane accounts for over two-thirds of the total energy or power required.

It should be kept in mind that the "friction" between tool and chip is also a shearing action, and in part of the region of contact the entire area undergoes shear in exactly the same manner as the shear plane does.

The fact that the condition of equilibrium is identical with an energy balance, if the workpiece distortion is neglected and the ideal plastic curve is assumed, leads to a rather queer and possibly significant result. For a given value of the force R on the chip there is no energy preference for any value of ϕ. All values of ϕ require the same sum of distortion energy and friction dissipation, although, of course, the shear stress on the shear plane will vary. The energy consideration appears only in deciding on the value of ϕ which will require the least value of R or H for a given shear yield stress (which may be a function of the normal stress). The first point of view presented by Ernst and Merchant[2] in which R, β, and α are assumed as given, or measured, quantities and the value of ϕ is found by differentiating the expression for the shearing stress τ_y and equating to zero does not seem to have any validity on the basis of energy. The second viewpoint presented by Merchant[5] in which the horizontal or cutting force is minimized (a least power or energy requirement), should take into account the variation of the "friction" angle β with ϕ, and this has not yet been done.

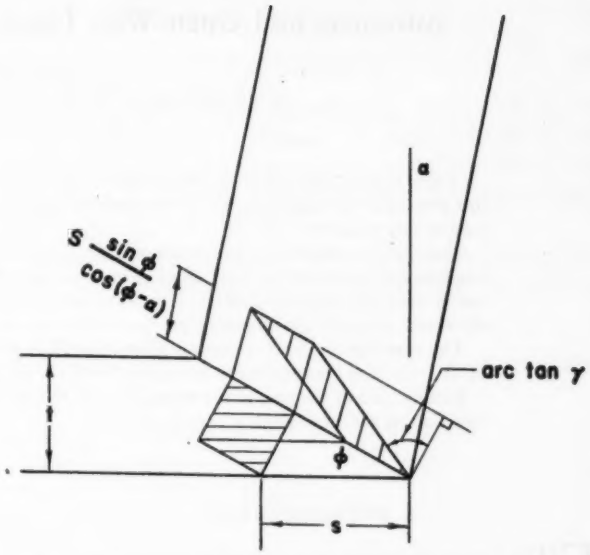


FIG. 14. Quantities required for energy calculations

$$\tan\gamma=\frac{\Delta V}{V'} \text{ (Fig. 3)}=\frac{\Delta V}{V_{chip}\cos(\phi-\alpha)}=\frac{\cos\alpha}{\sin\phi\cos(\phi-\alpha)}.$$

CONCLUSION

The analysis of the mechanics of metal cutting presented is of necessity qualitative because the basic information on the plasticity of metals is not as yet available. Very much more remains to be done before accurate predictions can be made for production processes. However, if inhomogeneity of metal is considered, the concept of time as the important variable permits the transition from the idealized picture of metal cutting to the practical.

It is interesting to speculate that, from the viewpoint of experimental plasticity, it may be metal cutting which will provide the information on the physical properties of metals and not this information obtained elsewhere which will enable the solution of the problems of metal cutting. There seems to be no other way to distort metal a large amount at enormous strain rates, without tremendous and unknown inertia forces, except by cutting. If so, the proper description of the mechanics of cutting will serve a twofold purpose, and experimental information is doubly valuable.

The author wishes to thank the Shell Oil Company, sponsor of a complete investigation of the fundamentals of metal cutting, for permission to publish this summary of a small part of the work, and to express his appreciation to Dr. H. Ekstein of Armour Research Foundation for the many helpful suggestions he has made.

Antennas and Open-Wire Lines. III. Image-Line Measurements*

PATRICK CONLEY**
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received December 20, 1948)

A new type of open-wire line is described for which many disadvantages, notably radiation and unbalance, are overcome by replacing one of the physical conductors by the image of the other in a semi-infinite, highly conducting plane.

The characteristics of a line driven at both ends and loaded at the center are examined; it is shown that the impedances measured on both sides of such a line are equal and simply related to the load impedance, provided that electrical symmetry is maintained about the load point. The double-ended drive requires no dielectric supports and permits the line conductor to continue past the load.

The construction and operation of an image line with double-ended drive is described. The characteristics of the new line compare very favorably with those of a good coaxial measuring line.

Finally, use of the line in the measurement of the apparent impedance of a cylindrical dipole antenna as a function of the line spacing is described.

1. INTRODUCTION

THIS paper describes the construction of an open-wire measuring line in which the disadvantages of radiation and stray couplings are largely overcome by the use of an inherently balanced system in which one of the line conductors is replaced by the image of the other in a semi-infinite, plane conductor. The properties of double-ended drive are examined for use in the measurement of the impedance of a cylindrical antenna as a function of the gap at the driving point. The advantages and theory of such a line are described in Section 2; descriptions of the apparatus itself are included in Section 3; while Section 4 describes the method of measurement and typical results.

The interpretation of the results and their correlation with theory is contained in Part I.[1]

2. ADVANTAGES AND THEORY OF AN IMAGE-LINE WITH DOUBLE DRIVE

If one of the conductors of an open-wire line is replaced by the image of the other in an infinite perfectly conducting plane, the resulting structure has all of the theoretical characteristics of the original two-wire line.

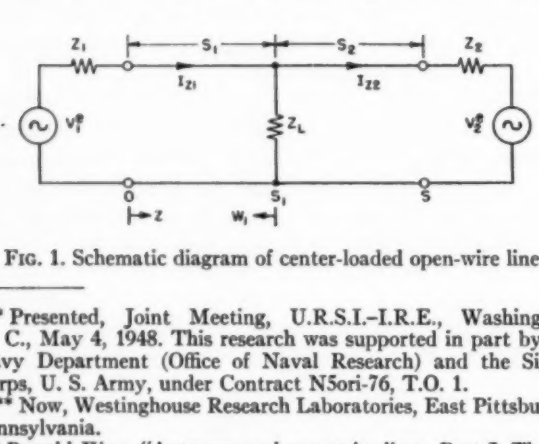


FIG. 1. Schematic diagram of center-loaded open-wire line.

The theory of images has been discussed at length in the literature[2,3] and need not be reviewed here.

An image line possesses two advantages: (1) It is inherently balanced. Any irregularities existing in or around the physical conductor are imaged exactly so that a high degree of electrical symmetry is readily obtainable. (2) The image plane provides an effective shield for the operator and measuring equipment.

Image planes simulating half of symmetrical structures have been used for some time in connection with antennas. The size of image plane necessary to approximate an infinite plane has been investigated, and planes four to six wave-lengths square have been found to give satisfactory experimental results.[4] A sufficient requirement is that an increase in size or a disturbance at the edge should have negligible effect on the observed quantities.

If an open-wire line of either conventional or image type is driven from both ends and loaded at the center, certain advantages are obtained: (1) The need for supports is reduced to those whose effects may be lumped with the internal impedances of the driving devices. (2) The line itself does not end abruptly at the load in the same sense that it does with single-ended drive. (3) The no-load termination may be regulated by the

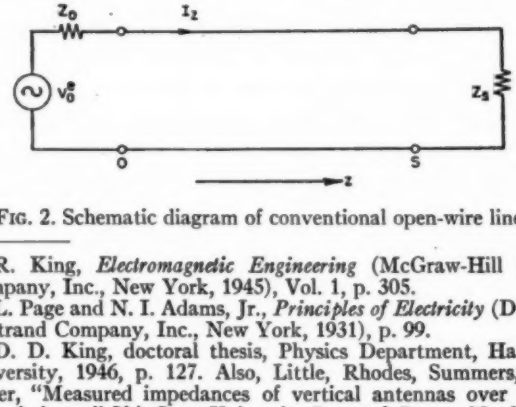


FIG. 2. Schematic diagram of conventional open-wire line.

* Presented, Joint Meeting, U.R.S.I.–I.R.E., Washington, D. C., May 4, 1948. This research was supported in part by the Navy Department (Office of Naval Research) and the Signal Corps, U. S. Army, under Contract N5ori-76, T.O. 1.

** Now, Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania.

[1] Ronold King, "Antennas and open-wire lines. Part I. Theory and summary of measurements," J. App. Phys. **20**, 832 (1949).

[2] R. King, *Electromagnetic Engineering* (McGraw-Hill Book Company, Inc., New York, 1945), Vol. 1, p. 305.

[3] L. Page and N. I. Adams, Jr., *Principles of Electricity* (D. Van Nostrand Company, Inc., New York, 1931), p. 99.

[4] D. D. King, doctoral thesis, Physics Department, Harvard University, 1946, p. 127. Also, Little, Rhodes, Summers, and Maier, "Measured impedances of vertical antennas over finite ground planes," Ohio State University Research Report No. 233–3, Antenna Laboratory, 1948.

relative magnitudes and phases of the driving currents, so that an ideal standard termination may be simulated.

A schematic diagram of the double-drive line is shown in Fig. 1. With reference to this figure, let

$$Z_i = Z_c \coth\theta_i, \ (i=1, 2, L, \text{etc.}), \tag{1}$$

$$\theta_i = \rho_i + j\Phi_i, \ (i=1, 2, L, \text{etc.}), \tag{2}$$

$$w_1 = s_1 - z, \tag{3}$$

$$s_2 = s - s_1 = s_1 + s_d, \tag{4}$$

$$V_2^e = \delta V_1^e, \ (\delta \text{ complex}), \tag{5}$$

$$\theta_2 = \theta_1 + \vartheta, \ (\vartheta \text{ complex}), \tag{6}$$

Z_c is the characteristic impedance, $\gamma = \alpha + j\beta$ is the complex propagation constant. I_{z1} is the current in the positive z direction at some point z_1 to the left of the load; I_{z2} is the current in the positive z direction at some point z_2 to the right of the load.

Using the foregoing definitions and the superposition principle, the total currents on the halves of the line are:

$$I_{z1} = \frac{V_1^e}{Z_c} \sinh\theta_1 \frac{G \coth\theta_L \sinh\gamma w_1 + \cosh\gamma w_1}{G \coth\theta_L \sinh(\gamma s_1 + \theta_1) + \cosh(\gamma s_1 + \theta_1)}, \tag{7a}$$

$$I_{z2} = \frac{-V_2^e}{Z_c} \sinh\theta_2 \frac{H \coth\theta_L \sinh(-\gamma w_1) + \cosh(-\gamma w_1)}{H \coth\theta_L \sinh(\gamma s_2 + \theta_2) + \cosh(\gamma s_2 + \theta_2)}, \tag{7b}$$

where

$$G \equiv \left[\frac{\cosh(\gamma s_1 + \theta_1 + \gamma s_d + \vartheta) + \delta \cosh\vartheta \cosh(\gamma s_1 + \theta_1) + \delta \coth\theta_1 \sinh\vartheta \cosh(\gamma s_1 + \theta_1)}{\cosh(\gamma s_1 + \theta_1 + \gamma s_d + \vartheta) + \coth\theta_L [\sinh(\gamma s_1 + \theta_1 + \gamma s_d + \vartheta) - \delta \cosh\vartheta \sinh(\gamma s_1 + \theta) - \delta \coth\theta_1 \sinh\vartheta \sinh(\gamma s_1 + \theta_1)]} \right], \tag{7c}$$

$$H \equiv \left[\frac{\cosh(\gamma s_2 + \theta_2 - \gamma s_d - \vartheta) + (1/\delta) \cosh(-\vartheta) \cosh(\gamma s_2 + \theta_2) + (1/\delta) \coth\theta_2 \sinh(-\vartheta) \cosh(\gamma s_2 + \theta_2)}{\cosh(\gamma s_2 + \theta_2 - \gamma s_d - \vartheta) + \coth\theta_L [\sinh(\gamma s_2 + \theta_2 - \gamma s_d - \vartheta) - (1/\delta) \cosh(-\vartheta) \sinh(\gamma s_2 + \theta_2) - (1/\delta) \coth\theta_2 \sinh(-\vartheta) \sinh(\gamma s_2 + \theta_2)]} \right]. \tag{7d}$$

Now consider the conventional transmission line in Fig. 2.[5] With $Z_0 = Z_c \coth\theta_0$, $Z_s = Z_c \coth\theta_s$, $w = s - z$, it follows that

$$I_z = \frac{V_0^e}{Z_c} \sinh\theta_0 \left[\frac{\coth\theta_s \sinh\gamma w + \cosh\gamma w}{\coth\theta_s \sinh(\gamma s + \theta_0) + \cosh(\gamma s + \theta_s)} \right]. \tag{8}$$

It is clear from (1) that $G \coth\theta_L$ is the normalized load impedance multiplied by the complex factor G. Comparison of (7) with (8) leads to the remarkable conclusion that the current—and thus the voltage—on the right and left sides of this line, driven at both ends and loaded in the center with an impedance Z_L, are exactly those obtaining on a conventional line, driven at one end and with load impedance Z_L/G or Z_L/H, depending on which side of the line the observation is made. The net effect of the double drive is to cause a given load impedance Z_L to be measured as GZ_L on the left side or as HZ_L on the right side. Impedance measurements on each side of the double-drive line may be made as if it were a conventional line, simply by remembering to divide the answer by G or H. This is complicated by the difficulty in determining G or H. In order to simplify the cumbersome expressions, let the line be the same on each side of the load. That is,

$$V_2^e = V_1^e, \quad (\delta = 1), \tag{9a}$$

$$\theta_2 = \theta_1, \quad (\vartheta = 0), \tag{9b}$$

$$s_2 = s_1, \quad (s_d = 0). \tag{9c}$$

With the assumptions (9) great simplifications may be made in (7c) and (7d). The result is

$$G = H = 2. \tag{10}$$

The practical problem is now simplified to measuring the impedance on either side of the double line and dividing the result by two. This assumes, of course, that the assumptions (9) can be fulfilled in practice.

Since it is unlikely that one could duplicate the halves of the line experimentally, an investigation has been made to determine the effect upon G and H caused by small departures from the conditions (9). By expanding (7) in a MacLaurin series it may be shown that the first-order effect of a departure from (9) is to cause G and H to vary equally and in opposite directions. Thus, it will be practically convenient and relatively accurate to keep the line as accurately symmetrical about the load as possible and to average the measurements on both sides of the line to eliminate residual errors.

The conditions of balance obviously are not entirely independent. A small error in line length, for instance, may be corrected partially by a slight change in the

[5] A discussion of the presentation of conventional line theory by use of terminal functions may be found in R. King, "Transmission line theory and its application," J. App. Phys. 14, 577 (1943).

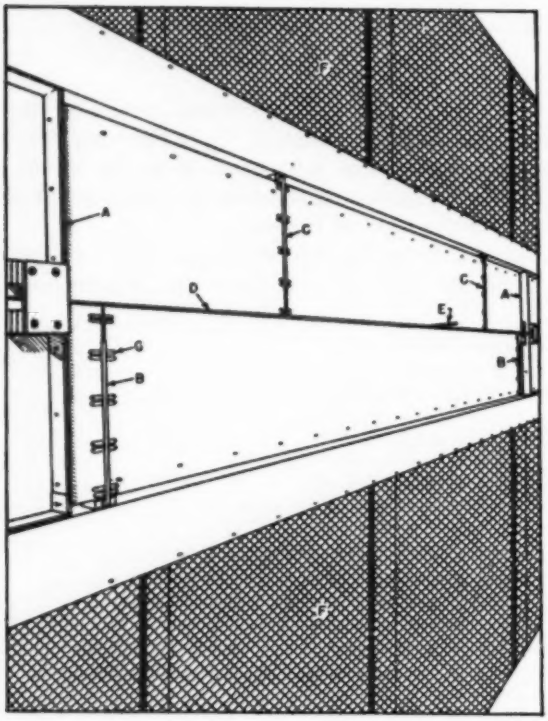


FIG. 3. Front view of the image line. The various components are identified as follows: short-circuiting bridge A; driver probe B; detector probe C; line conductor D; antenna E; galvanized screen F; phosphor bronze fingers G.

phase of the applied voltage, etc. If the line is electrically symmetrical there should be a current null at the load; that is, the currents from both ends should be exactly equal in magnitude and opposite in direction at that point. This is a convenient check for symmetry. Another condition resulting from symmetry is that the current (and hence the voltage) distributions should be symmetrical on both sides of the load. This results in equality of impedance measurements on the two sides, and such equality is a check for balance.

An image line with double drive provides a good solution to the principal problems that have led, in the past, to a preference for coaxial lines over open-wire lines in measurement work. The construction of such a line is described in the next section.

3. APPARATUS

The image plane (Fig. 3) is in three sections. The center, located immediately behind the line conductor, is a solid sheet of $\frac{1}{8}$-inch aluminum ST-24; while the top and bottom sections are $\frac{5}{16}$-inch mesh, galvanized screen, bound at the central edges with 5-inch strips of $\frac{1}{8}$-inch aluminum ST-24. Between these strips and the center sheet is a $\frac{3}{4}$-inch slot for the insertion of detectors, driving units, etc. The over-all useful size of the image plane is 12 feet $\times$10 feet, or approximately 9.15$\times$7.65 wave-lengths at the operating frequency of 750 megacycles per second. This is adequate to image an antenna load at the center; for the line, the aluminum sheet alone would probably suffice, since the close spacing assures that the greater part of the image effect is due to currents in the plane near the line, and the effects of conductor current and image-plane current cancel at appreciable distances. The lower conductivity of the outer sections is tolerable since the current densities there are greatly reduced. The effect of the slots has been examined on a scale model[6] at 3000 megacycles per second, and results show that the surface-current distribution is not altered by their presence. The slots in the region near the center were covered and no change was observed in the measured impedance.

The single line conductor is supported under tension by adjustable fittings which permit variations of from $\frac{1}{32}$ inch to two inches in clearance between conductor and plane, as well as vertical adjustments of conductor position. The sag in the line conductor can be reduced to a negligible amount.

The bridges at both ends are movable and are designed to provide effective short circuits for the line. The line is connected electrically to the image plane by pairs of phosphor bronze fingers separated by a quarter-wave-length. In addition, a choke section reduces propagation along the image plane past the bridges. One of the short-circuiting bridges with its associated driver probe is shown in Fig. 4.

The mechanism for moving and locating the short-circuiting bridges, driving units, and detectors is attached to the rear of the image plane. The various units are mounted on brass bearings which travel on steel rails. They are moved by cables, pulleys, and counterweights. Provision is made for the following operations: (1) All units can be moved by hand independently; (2) one detector may be fixed and the other moved mechanically; (3) both detectors may be moved mechanically and simultaneously in such a way as to maintain equal distances to the point midway between them (usually chosen as the load in the double-drive arrangement); (4) one short-circuiting bridge may be fixed and the other moved mechanically; (5) both bridges may be moved mechanically and simultaneously to maintain equal distances to the point midway between them; and (6) the driving units may be attached at any desired distance from the bridges and move with the bridges or instead of the bridges.

The driver and detector probes consist of telescoping sections of coaxial lines with inner conductors that extend beyond the shields and outer conductors that are grounded to the image plane by sliding contacts.

The manner in which the image line is driven is indicated by the block diagram in Fig. 5. The oscillator is of conventional type employing a "butterfly" tank circuit. Energy is extracted from the tank by means of three coupling loops, two adjustable from the front panel and connected to the driving units. The third loop is ex-

[6] D. K. Reynolds, unpublished research, Cruft Laboratory, Harvard University, 1947.

tremely loosely coupled; it feeds a power monitoring crystal and a frequency monitoring wave meter. Variable relative phasing between the two driving units is controlled by a line-stretcher in series with one of the feed lines. Numerous low pass filters effectively eliminate harmonics in the output; their importance cannot be overemphasized.

The radiofrequency energy picked up by the detector probes is fed through a double-stub matching section into the detecting element. Although any suitable detector may be employed, the present equipment is provided with a cartridge in which may be mounted a series 1N21B crystal or a series Littelfuse bolometer element.

The load is normally connected to the center of the line. If the characteristic impedance of the transmission line is taken to be that of the two-wire line, without image plane, the measured load impedance is twice that of the impedor terminating the line. If the line is operated with double drive so that $G=H=2$, the measured impedance is apparently four times the value of the impedance actually present. In antenna measurements, the presence of the image antenna element is highly convenient in the interests of symmetry. This may or may not be the case with other impedance elements.

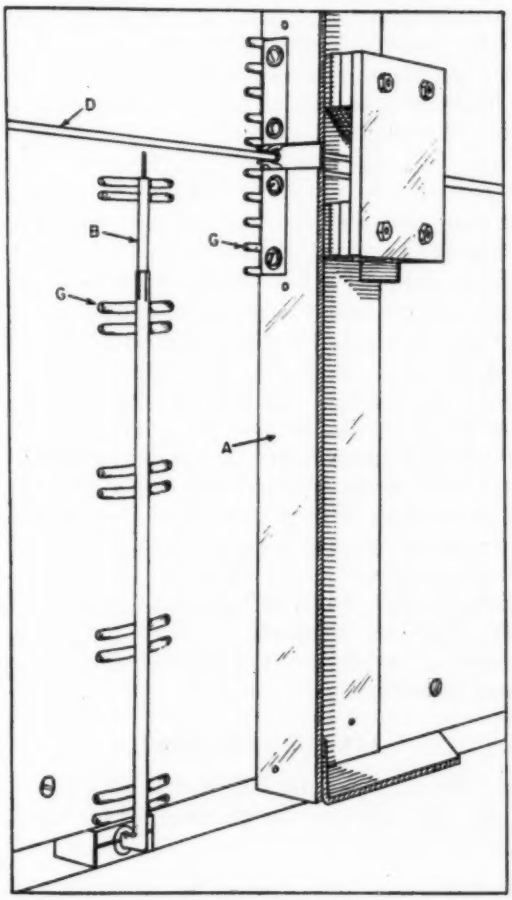


FIG. 4. Detailed view of the driving probe and short-circuiting bridge.

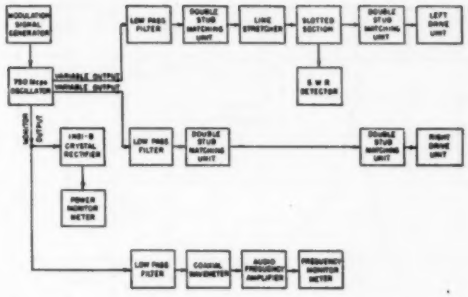


FIG. 5. Block diagram of the complete generator and monitor for double-ended drive.

4. METHOD OF MEASUREMENT

The natural reference termination of the line, used with double drive, is the open circuit. This is obtained easily in practice because in normal operation with no load the line currents must vanish at the center. This current null is equivalent to an open circuit at the center of the line.

The adjustment of the phase and magnitude of the voltages applied to the two ends of the line so as to produce a current zero at the center of the line is difficult. The method used relies upon the requirement of similarity in the measurements made on the two sides of the line. The trimming procedure follows: (1) With power fed to both sides of the line and all components adjusted for mechanical symmetry about the center of the line, the line stretcher is adjusted until voltage minima are equidistant from the load on both sides. (2) The power supplied to one or both sides of the line is varied (by means of the adjustable oscillator coupling loops) until the standing wave ratios are equal on both sides. (3) The entire procedure is repeated until the distributions of voltage on both sides of the line are identical, or very nearly so. With no load, detectors loosely coupled, and the line properly trimmed, the power standing wave ratio is of the order of 10^4. This compares favorably with that obtainable on good coaxial lines at the same frequency.

Investigations into the theoretical conditions at the driving point of a dipole antenna excited with an open-wire line have King[1] and Winternitz[7] to the conclusion that there is a coupling between the antenna and the line which causes a change in the line constants near the load as a function of the line spacing. Since the effect of this coupling is approximately twice as great for a center-loaded line using double-ended drive as for an end-loaded line, it is an appropriate subject for experimental investigation using the apparatus just described. The experimental problem was to measure the impedance of a dipole antenna on the image line for several values of line spacing. The antenna whose impedance was measured was identical to the single line conductor in

[7] T. Winternitz, doctoral thesis, Department of Engineering Sciences and Applied Physics, Harvard University, 1948. Accepted for publication in Quarterly of Applied Mathematics.

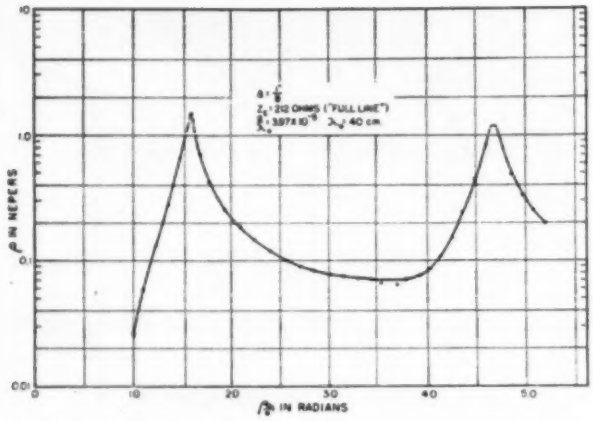


FIG. 6. Damping function of dipole antenna as a function of antenna half-length for "full line" spacing of 0.0238λ.

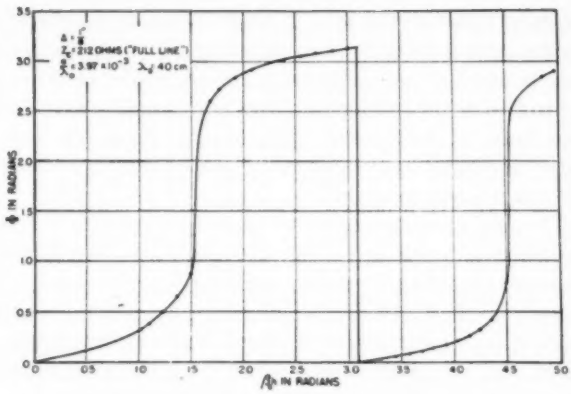


FIG. 7. Phase function of dipole antenna as a function of antenna half-length for "full line" spacing of 0.0238λ.

material and diameter ($\frac{1}{8}$-inch O.D. brass tubing). The junction of the line and antenna was carefully machined. The outer end was fitted with a hemispherical cap that could be removed in order to cut off the antenna. Data were taken for "full-line" spacings of 0.0159λ, 0.0238λ, 0.0556λ, and 0.1032λ (measured from center of physical conductor to center of imaginary image conductor); corresponding to values of Δ (clearance between image plane and line conductor); of $\frac{1}{16}$ inch, $\frac{1}{8}$ inch, $\frac{3}{8}$ inch and $\frac{3}{4}$ inch respectively. Both the standing-wave ratio method and the distribution-curve-dip-width method were used for the determination of damping.[8] To determine the phase function of the termination, the distribution-curve minimum was carefully plotted in every case. Data from both sides of the line were averaged to give the result of each measurement. The antenna was shortened between successive measurements by cutting off and facing on a lathe, and the line trimming was adjusted with the antenna in place at the beginning of each measurement. Figures 6 and 7 show typical curves of the experimentally determined damping and phase functions ρ and Φ. The impedance is computed from the following relations.[9] Impedance curves are in Fig. 25a of Part I.

$$2R=\frac{R_c \sinh 2\rho}{\cosh 2\rho-\cos 2\Phi}, \tag{19}$$

$$2X=\frac{-R_c \sin 2\Phi}{\cosh 2\rho-\cos 2\Phi}. \tag{20}$$

The principal difficulties in measurement were a result of the low oscillator power available. The close coupling between the driving probes and the transmission line required the line to be trimmed for each length of antenna. As the procedure used was tedious, retrimming between measurements was a disadvantage. The line was found exceedingly difficult to trim when the antenna was near its resonant length. Usually the standing wave ratio was small, and the accurate location of minima was impossible. For this reason the measured values of the phase and damping functions are least accurate in the region near resonance. For very short antenna lengths, the detector readings at the minima are very small. For antenna lengths below about one radian, the readings at the minima begin to be affected by crystal and amplifier noise. The result is a small error in the value of damping.

A very small outward bow in the center of the image plane proved to be a problem. Actually, it causes a very small change in the line spacing over a considerable distance. With no load, the effect is similar to that of a small capacitance located at the center of the line the reactance of which may be measured using conventional techniques. While the effect is insignificant when the line spacing is large, this is not true with very small line spacings. An attempt was made to correct the impedances by assuming them to be in parallel with the measured capacitance lumped at the driving point. Since this is not an accurate correction, but at best an estimate of its order of magnitude, both "corrected" and uncorrected values are shown in Fig. 25b of Part I. The two "corrected" sets of curves in Fig. 25a in Part I are still not actually correct, as is directly apparent. Nevertheless, they are adequate to show the interesting continuity in the effect of line spacing on the apparent terminal impedance.

ACKNOWLEDGMENT

The writer is especially indebted to Professor R. W. P. King who suggested the problem for his invaluable advice.

[8] D. D. King, "Impedance measurement on transmission lines," Proc. I.R.E. **35**, 509–514 (1947).

[9] D. D. King and R. King, "Terminal functions for antennas," J. App. Phys. **15**, 186–192 (1944).

Convection Currents in Porous Media
II. Observation of Conditions at Onset of Convection

H. L. MORRISON,
North Carolina State College of Engineering, Raleigh, North Carolina

F. T. ROGERS, JR.,*
The University of North Carolina, Chapel Hill, North Carolina

AND

C. W. HORTON
Defense Research Laboratory, The University of Texas, Austin, Texas

(Received January 31, 1949)

Small-scale experiments have been carried out for determining the minimum vertical thermal gradient which is required to cause convection in liquids entrapped in porous media. Observations relative to the onset of convective flow in unconsolidated sands indicate that the present theories predict minimum gradients which are excessive by considerable amounts, possibly because they neglect the temperature-dependence of viscosity. The ratio of theoretical to observed gradients is found to be roughly

$$R=(Kh^2\mu_{Av}/k\rho\alpha)^{0.51},$$

where h^2 is the thermal diffusivity, μ_{Av} is average viscosity, k is flow-permeability, ρ is density of liquid, α is the coefficient of cubical expansion of the liquid, and where $K=10^{-3}$ sec.^{2} per cm^2 °C for c.g.s. units. By extrapolation, it is possible to strengthen the earlier conclusion that convection occurs in the Woodbine sand near the Mexia fault zone.

IN an earlier paper of the same general title,[1] a theory was presented for the minimum thermal gradient (β_c) required to cause the vertical convective flow of a liquid through a porous medium; according to this theory, which made no allowance for the variation of viscosity of liquid with temperature in the medium,

$$\beta_{c,\,th}=4\pi^2h^2\mu/k\rho_0g\alpha D^2, \qquad (1)$$

where μ, α, and ρ_0 are the viscosity, coefficient of cubical thermal expansion, and density of the liquid respectively, where k and D are the flow-permeability and vertical thickness of the porous medium, and where h^2 is the thermal diffusivity of the liquid-and-medium mixture; g is the acceleration of gravity. Preliminary laboratory measurements have been made in an attempt to verify Eq. (1); these, as previously reported,[2] seemed to indicate significant limitations to the accuracy of the theory and thereby led to the present more extensive experimental test of Eq. (1).

OBSERVATIONS

All of the experiments here reported were made on unconsolidated media. In one instance steel filings were used as the medium, but in all others, screened sand from White Lake (Bladen County, North Carolina) was used; this sand has a very light grey appearance, almost white, and is generally free of dust-sized particles.

The permeability of each test-medium was obtained from observations of the downward flow of water through a column of medium, by the general method of Terzaghi.[3] If a uniform vertical tube containing depth L of medium, be filled with water to a depth d, where $d>L$, then the downward flow of water through the medium under the action of gravity is described by

$$\ln(d/d_0)=-(k\rho g/\mu L)\tau; \qquad (2)$$

here d_0 is the value of d at time $\tau=0$, and the other symbols have the meanings specified in connection with Eq. (1). Provided suitable allowances be made for compaction of the medium during successive sets of

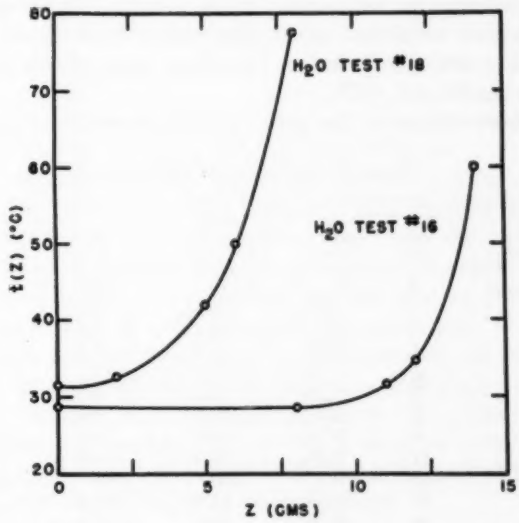


FIG. 1. Typical vertical variation of temperature.

* Present address: U. S. Naval Ordnance Test Station, Inyokern, California.

[1] C. W. Horton and F. T. Rogers, Jr., J. App. Phys. **16**, 367 (1945). E. R. Lapwood has extended these calculations considerably; see reference 7.

[2] H. L. Morrison, J. App. Phys. **18**, 849 (1947).

[3] See M. Muskat, *The Flow of Homogeneous Fluids through Porous Media* (McGraw-Hill Book Company, Inc., New York, 1937), p. 84.

TABLE I. Conditions at onset of convection.

Liquid	No. of test	Medium	k in darcies $\times 10^3$	z_0 (cm)	D (cm)	Horizontal extent of medium (cm)	Temperatures, °C $t(0)$	$t(z_0)$	$t(D)$	t_{AV}	Observed β_e Gradient (°C/cm) Local	Mean
Glycerol	2	Sand	97	2.5	10	7	133	70	37	58.9	81	9.4
CCl_4	3	Sand	127	2.5	10	7	30.8	28.8	22.8	26.8	0.8	0.8
H_2O	1	Sand	125	2.5	10	7	63	38.2	25	34.6	6.4	3.8
H_2O	5	Sand	70	2.5	10	7	95	60	34	51.8	11.2	6.1
H_2O	6	Sand	88	2.5	10	7	74	40	20	33.5	8.5	5.4
H_2O	7	Sand	125	2.5	10	15	65	38	25.5	34.5	6.4	3.95
H_2O	8	Sand	125	2.5	10	4	65	36	24	33	6.24	4.1
H_2O	9	Sand	125	2.5	14	7	60	32.5	29	32	3.2	2.2
H_2O	11	Sand	125	2.5	8	7	90	46	35	45.6	10.0	6.9
H_2O	13	Steel	146	2.5	10	7	100	—	<25	—	—	>7.5
H_2O	15	Sand	125	5	10	15	69	34	31.5	38.7	1.8	3.75
H_2O	16	Sand	125	7	14	15	60	29.1	29	32	0.2	2.2
H_2O	18	Sand	125	5	8	15	77.5	34.4	32.0	44.5	2.4	5.6

observations, k can be got conveniently from a semi-logarithmic graph of d/d_0 *vs* τ; the significance of k-values got in this way has been reported elsewhere.[4]

In each experiment a glass vessel filled to depth D with a mixture of medium and liquid, was heated from below by means of electric heating coils placed at some distance from the bottom of the vessel. Temperatures in the first few experiments were measured with iron-constantan thermocouples; later they were measured more conveniently by mercury-in-glass thermometers, after careful measurements had shown that such thermometers did not interfere appreciably with the convective flow. The horizontal uniformity of temperature reached in this simple thermal system was not perfect; but $\pm\frac{1}{4}$°C variations from uniformity were the largest encountered. In every case heat was applied continually, leading to a rise of temperature of, and in temperature gradient in, the mixture until the onset of convection was observed. Typical vertical variations of temperature $t(z)$, at the onset of convection, are shown in Fig. 1; here z is the distance from the bottom of the vessel. It is noteworthy that while many of the vertical temperature-variations exhibited great departures from linearity, much more nearly linear variations were obtained in tests made with CCl_4.

Observations of the onset of convection were made through the wall of the glass vessel by means of a microscope. The initial motion of the liquid was shown in some instances by the motion of occasional tiny suspended particles of impurity. More commonly a layer of soluble dye was inserted in the mixture to indicate this initial motion; Congo red was used in the water mixtures, and Sudan IV in the CCl_4 mixtures. The initial motion, at the rather slow rate of about 0.001 cm/sec., always occurred rather abruptly in the present tests; therefore the data relative to the onset of convection apply more precisely to the attainment of such rates of flow as this. Repeated trials, however, showed that a condition of no motion (0.00001 cm/sec. or less) persisted to temperature-distributions which were experimentally indistinguishable from those appropriate to the 0.001 cm/sec. flow, so that the observed slow-flow condition is practically the same as the onset condition.

A number of mixtures were examined by excavation after convection had continued for some time, in order to determine the character of migration of the dyes. Invariably the upper boundary of dyed liquid appeared to have moved upward essentially bodily; if there were regions of differing directions or amounts of migration, they were either too small to be noticed in the surfaces exposed by excavation or were obliterated in that surface by the process of excavation. There was thus no evidence that the liquid at the walls of the vessel partook of convection to a greater or less extent than the liquid deep inside of the mixture. Subsidiary tests were made relative to the diffusion of the dyes; only in the case of CCl_4 was diffusion appreciable, and there the rate was about 0.0001 cm/sec.; the onset of convection in the present experiments was therefore not complicated to any great extent by the effects of diffusion.

In an attempt to study most of the dependences required by Eq. (1), experiments were made with several depths (D) of mixture, with several permeabilities of medium, with different horizontal extents of mixture, at different temperatures, and using different liquids. Observations of the onset of convection were made at several distances z_0 from the bottom of the vessel, to obtain information as to the extent of convective flow at the onset. In no respect however, can the ranges over

TABLE II. Reduction of the data.

Liquid	No. of test	μ_{AV} (centi-poises)	β_e by Eq. (1) (°C/cm)	R	β_e by Eqs. (3) (°C/cm)
Glycerol	2	40	550	58	14
CCl_4	3	0.88	2.86	3.6	1.1
H_2O	1	0.73	26.3	6.9	3.1
H_2O	5	0.53	34.1	5.6	3.5
H_2O	6	0.74_5	38.4	7.1	3.7
H_2O	7	0.73_1	26.3	6.7	3.1
H_2O	8	0.75_2	27.0	6.6	3.1
H_2O	9	0.76_8	14.1	6.4	1.6
H_2O	11	0.59_2	33.3	4.8	4.4
H_2O	13	0.3	460	<61	<12
H_2O	15	0.67_5	24.3	6.5	3.0
H_2O	16	0.76_8	14.1	6.4	1.6
H_2O	18	0.62_3	40.2	7.2	4.7

[4] H. L. Morrison and F. T. Rogers, Jr., Phys. Rev. **71**, 834 (1947).

which the variables extend be considered as a truly definitive assessment of Eq. (1); on the other hand it is possible that Eq. (1) should be refined somewhat before an exhaustive experimental study is undertaken.

Table I contains typical data concerning conditions at the onset of convection. The data relative to temperature gradients were obtained from charts like Fig. 1; thus $(dt/dz)_{z_0}$ is the local gradient at the point, z_0, of observation, while $[t(D)-t(0)]/D$ is the mean gradient which should be comparable with the β_c predicted by Eq. (1); the average temperature is

$$t_{AV}=\int_0^D t(z)dz \Big/ \int_0^D dz.$$

Significantly it was not possible to produce convection in the porous steel medium, probably because of the high thermal diffusivity of the steel-and-water mixture.

DISCUSSION

It can be seen directly from Table I that the thermal gradients required to initiate convection depend upon the several factors in Eq. (1) qualitatively correctly. Moreover, as may be seen by Tests Nos. 1, 7, and 8 with H_2O, there is no evident dependence of required gradient upon the horizontal extent of the medium. A further interesting aspect of the initial convective-flow process is illustrated by Tests Nos. 9 and 16 with H_2O: whereas the local gradients at z_0 are about in the ratio 16:1, the mean gradients are the same. This latter situation, also shown less strikingly by Tests Nos. 7 and 15 with H_2O, seems to indicate that convection occurs at the onset quite generally throughout the medium and does not, in spite of local variations of gradient, occur merely locally.

A more nearly precise comparison of the data with Eq. (1) is difficult to make, however, because of the temperature-dependence of viscosity; for example, the wide difference between $t(0)$ and $t(D)$ in the case of glycerol implies that the local viscosity varied by about an order of magnitude between the top and the bottom of the vessel. Since the theory from which Eq. (1) derives, does not contemplate this temperature-dependence of viscosity, it is not clear what value of μ should be chosen for any particular comparison. If the general convective motion be thought of as governed by the average (throughout the liquid) viscosity, so that a μ_M-value can be specified in accord with t_M, then a comparison of sorts can be made. Column 4 of Table II shows the gradients $(\beta_{c,\,th})$ calculated by Eq. (1) using $\mu=\mu_M$, while column 5 shows the ratios (R) of these to the observed mean gradients $(\beta_{c,\,obs})$.

In computing column 4 of Table II, h^2 was taken as 0.002 cm²/sec. for the several sand-mixtures, and as 0.1 cm²/sec. for the iron-mixture. The values 0.475, 1.206, and 0.177 relative to sand were used for $1000\times\alpha$, while 1.26, 1.595, and 1.00 were used for ρ_0; these refer to glycerol, CCl_4, and H_2O respectively. Like the values used in Table II for μ_M, these are accepted c.g.s.-values taken from standard tabulations of the physical properties of matter.

Clearly, as was anticipated in reference 1, $R>1$. That is, Eq. (1) predicts values of β_c, using $\mu=\mu_M$, which are excessive by a factor which may be anything between 1 and 100, depending upon the physical situation. It is worth mentioning that $\beta_{c,\,obs}$ could be expected to be less than $\beta_{c,\,th}$ in the present tests if upward circulation should serve to separate the solid particles and so produce larger permeabilities; this separation should not occur prior to the onset of convection, and indeed, no displacements of particles were visible in microscopic observations during the onset of convection. It is also noteworthy that a more careful allowance for boundary conditions does not appreciably affect R; according to the calculations of both Low[5] and Lapwood,[6] the former for free fluids and the latter for fluids in porous media, the right member of Eq. (1) should be slightly increased by a constant factor, to take account of the fact that the lower surface of fluid in the present experiments was "rigid." Since this factor amounts to something like 1.3, it is hardly significant.

The great range of values shown for R in Table II suggests that R might well be correlated with one or more of the variables appearing in Eq. (1). It is not clear just what correlation would be the most significant physically, but R can be related simply to the quantity $h^2\mu_M/k\rho\alpha$, which is characteristic of the medium and its entrapped fluid, but not of the dimensions of the extent of fluid and medium. Thus

$$\left.\begin{aligned}\beta_{c,\,obs}\cong\beta_{c,\,th}(k\rho\alpha/h^2\mu_M K)^{0.51},\\ 7\leqslant Kh^2\mu_M/k\rho\alpha\leqslant 2000,\end{aligned}\right\} \quad (3)$$

where $\beta_{c,\,th}$ is given by Eq. (1) with $\mu=\mu_M$ and where $K=10^{-3}$ sec.²/cm²°C for c.g.s. units. The gradients calculated by Eqs. (3) are shown in the last column of Table II, and should be compared with the experimentally observed gradients in the last column of Table I. While better correlations can be had by more complicated formulas, the present data hardly justify their use.

The present results, when examined in the form of an R-*vs.*-$Kh^2\mu_M/k\rho\alpha$-chart, do not exhibit any clear correlation with D. This may indicate that the D^2-factor in Eq. (1) is correct, or that the data are too few to indicate a "scale"-type departure[7] from the D^2-factor.

Extrapolated to the geophysical problem of the existence of convection in the Woodbine sand near the Mexia fault zone, Eq. (3) leads to the result that the existing gradient is some 50 times as great as that required for convection; Eq. (1) indicates that the existing gradient is only 0.05 of that required for convection. Clearly neither the factor of 50 nor the much smaller one of 0.05 can be expected to be really accurate. Yet the evidence now weighs more strongly in favor of, rather than against, convective flow in this geological formation, and so is in better accord with the independent information[8] on the diffusion of NaCl there.

[5] A. R. Low, Proc. Roy. Soc. **A125**, 181 (1929).
[6] E. R. Lapwood, Proc. Camb. Phil. Soc. **44**, 508 (1948).
[7] M. K. Hubbert, Bull. Geol. Soc. Am. **48**, 1459 (1937).
[8] C. W. Horton, Am. Ass. Petrol. Geol. Bull. **28** (No. 11), 1635 (1944).

A Direct Method of Determining Preferred Orientation of a Flat Reflection Sample Using a Geiger Counter X-Ray Spectrometer

L. G. SCHULZ
Institute for the Study of Metals, University of Chicago, Chicago, Illinois
(Received March 21, 1949)

A new reflection method is described for determining pole figures of flat samples with a Geiger counter x-ray spectrometer. The chief advantage of the method is that the experimental data may be used directly without corrections for changes in geometry during rotation of the sample. Beginning from an initial position identical with the usual reflection arrangement the sample is rotated about an axis defined by the intersection of the sample surface with the plane of the spectrometer. During this motion a randomly oriented sample of sufficient thickness will yield a constant counting rate because the absorption and effective scattering volume of the sample remain unchanged. As a result, no correction formula is required. The experimental arrangement requires three horizontal slits; two are used to collimate the incident beam, and the third is placed in front of the counter. A mathematical analysis of the optical elements of the arrangement is given together with the results from a practical application.

INTRODUCTION

RECENTLY, several methods have been described for the determination of pole figures with a Geiger counter x-ray spectrometer.[1,2] The transmission method of Decker, Asp, and Harker employs a thin sheet specimen which can be rotated about two mutually perpendicular axes, one axis being normal to the plane of the sample and the other coinciding with the axis of the spectrometer. A factor is derived for correcting observed counting rates for changes in absorption and scattering volume as the sample is rotated. In the method of Norton, small rods of the sample material are used. These are mounted on the axis of the spectrometer and rotated. Because there is no variation due to changes in absorption or scattering volume, no correction formula is required.

In this paper a new reflection method will be presented in which a flat sample of sufficient thickness is rotated in such a manner that absorption and effective volume of scattering material remain constant. As a result, there is no change in counting rate during the rotation of a sample possessing random crystal orientation. The method, therefore, has the advantage of Norton in that no correction formula is required, and in addition, has the further advantage that samples of special shape are not necessary. This reflection method is characterized by the use of three horizontal slits, two serving to collimate the incident beam and the third being used at the entrance of the counter. During examination the sample is rotated about two axes, one normal to the sample, and the other located along the intersection of the sample and the plane of the spectrometer.

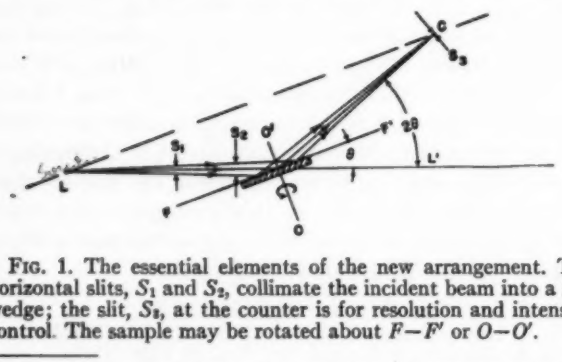


FIG. 1. The essential elements of the new arrangement. The horizontal slits, S_1 and S_2, collimate the incident beam into a flat wedge; the slit, S_3, at the counter is for resolution and intensity control. The sample may be rotated about $F-F'$ or $O-O'$.

PRINCIPLES OF THE METHOD

Using Fig. 1, a description will be given of the usual arrangement for a reflection sample in a Norelco instrument. L is a vertical line source of x-rays not more than 0.005 in. in width and about 0.050 in. long. From it diverges a wedge-shaped beam whose angular width is controlled by interchangeable vertical slits at the position S_1. This beam strikes the flat sample set at the Bragg angle θ. The diffracted beam leaving the sample converges to the vertical slit S_3 at the entrance of the counter C. This is the standard focusing arrangement for a flat reflection sample having random crystal orientation. The arrangement is characterized by vertical elements; the source is vertical as are also the slits S_1 and S_3. The mechanics of the instrument are such that the sample turns through one-half the angle through which the counter is rotated.

The question now is this: If the counter is kept at 2θ, how might the sample be rotated so as to determine the plane density of a sample having not random, but preferred orientation. It is desired that there be no correc-

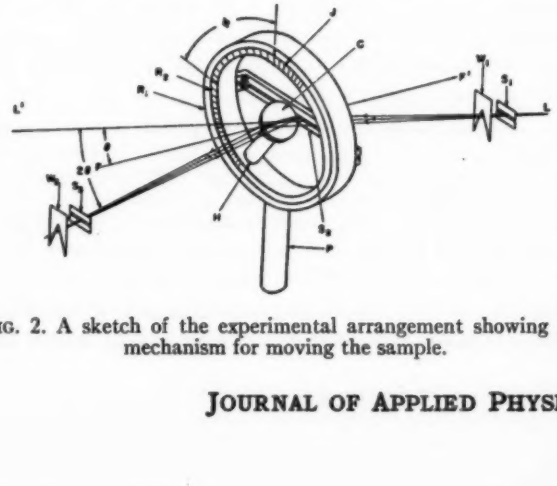


FIG. 2. A sketch of the experimental arrangement showing the mechanism for moving the sample.

[1] Decker, Asp, and Harker, J. App. Phys. **19**, 388–392 (1948).
[2] J. T. Norton, J. App. Phys. **19**, 1176–78 (1948).

tions for geometric changes, which is equivalent to requiring constant counting rate during the rotation of a sample having random orientation. Of equal importance is the desire to retain most of the focusing properties of the usual reflection arrangement.

What proved to be the best procedure for satisfying these requirements was found partly by experiment and partly from optical considerations. Several modifications of the usual experimental arrangement were necessary: (1) Narrow *horizontal* slits were placed at S_1 and S_2 to collimate the incident beam into approximate parallelism in the vertical direction but still diverging in the horizontal plane. At the entrance of the counter was placed a third horizontal slit, S_3. S_1 and S_2 were about 0.020 in. in width, and S_3 was varied from 0.020 in. to 0.050 in. Wedges were used to control the length of S_1 and S_3, but S_2 was unrestricted in length. (2) A mechanical mounting for the sample was added which permitted rotation about the axes $F-F'$ and $O-O'$.

These changes are more easily seen in the actual experimental arrangement given in Fig. 2. The post, P, fits into the standard sample holder of the Norelco instrument. This post supports an outer ring, R_1, in which rotates a smaller inner ring, R_2, the extent of the rotation, ϕ, being indicated by a scale, J. A second post, H, supports the sample, C. H is adjustable in a manner which permits the surface of C to be placed normal to the axis of H and also to contain the ring axis $F-F'$. This adjustment is maintained as the sample is rotated in its own plane. The collimating system, $S_1-S_2-W_1$ permits only a strip of the sample lying near the axis $F-F'$ to be irradiated. S_2 is defined by two long metal strips fastened to the ring R_1; the other slits and wedges are those included with the Norelco instrument. S_3 is normal to the Debye-Scherrer rings as it should be to gain resolving power since it is the variations in the intensity along the rings that reveals the preferred orientation. As the sample is turned about $F-F'$ defocusing produces a slight broadening of the rings, but this is not objectionable since S_3 is long enough to cover the entire ring width.

X-RAY REFLECTION FROM A FLAT SAMPLE NOT IN THE TRUE FOCUSING POSITION

Depending on the point of view, two explanations are possible for the constant counting rate as ϕ is changed. The first, and non-mathematical one, will be given at this point while the second, which is mathematical, will be delayed until the Appendix. The second treatment is more general in application and can be extended to the case of thin samples.

From the non-mathematical point of view there is no change in absorption or scattering volume during rotation. This will be demonstrated with the help of Fig. 3. In A the sample is shown as seen along $F-F'$ with ϕ equal to zero. Imaginary horizontal planes divide the sample into lamellae of infinitesimal thickness. The effect of rotation is equivalent to displacing these lamellae in the manner shown in B. Each lamella still retains the same effective scattering volume, and since the angle of incidence of the incident x-ray beam is unchanged, the absorption is the same as in A. Consequently, the diffracted intensity is independent of ϕ. There is some displacement of the diffracted beam accompanying the displacement of the lamellae but because the collimating system employs horizontal slits, there is no change in the energy entering the counter. The argument applies equally well to samples having

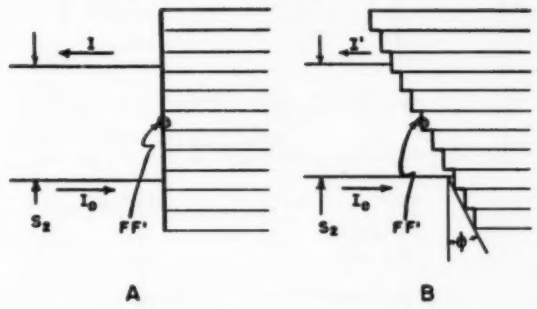


FIG. 3. Diagrams used to demonstrate that for a sufficiently thick sample there is no change in absorption or scattering volume during rotation about the axis $F-F'$.

preferred orientation as to those having random structure. The only requirement is that the sample have a thickness sufficient to give total absorption.

EXPERIMENTAL VERIFICATION AND APPLICATION

Although the argument just given applies for all values of ϕ up to 90°, in practice this limit cannot be reached because of imperfections in the alignment of the elements of the system and because practical considerations demand slits of finite width. A very convenient experimental test of the usable range of ϕ was through the use of samples known to have random orientation. For this purpose powder of any of the common insoluble halides such as CaF_2, MgF_2, and LiF was found satisfactory. The powdered material was pressed into a conventional sample holder and the surface smoothed in the usual way. A representative result for LiF using the (111) reflection is shown in Fig. 4, curve A. It was found that at about 75° there began a slight falling off the counting rate and beyond 80° the deviation became very great. These limits could be extended somewhat by decreasing the widths of the collimating slits S_1 and S_2.

An application to a typical problem in metallurgy was the determination of the orientation in rolled copper. After a 97 percent reduction a sample gave curve B of Fig. 4. This curve shows the counting rate as a function of ϕ when using the (111) reflection. It corresponds to the variation in intensity in the associated pole figure along the radius pointing in the rolling direction. The result is in agreement with that given by photographic methods. The last two curves, C and D, of Fig. 4 are included to show the high resolving power of the method. When copper recrystallizes (100) planes are parallel to

the surface. Using the (100) reflection curve C shows the perfection of the orientation of the recrystallized component in the direction transverse to the rolling; D gives the same information in the rolling direction. The width at half-maximum in C is only one degree, in fact, the peak resembles that given by a single crystal.

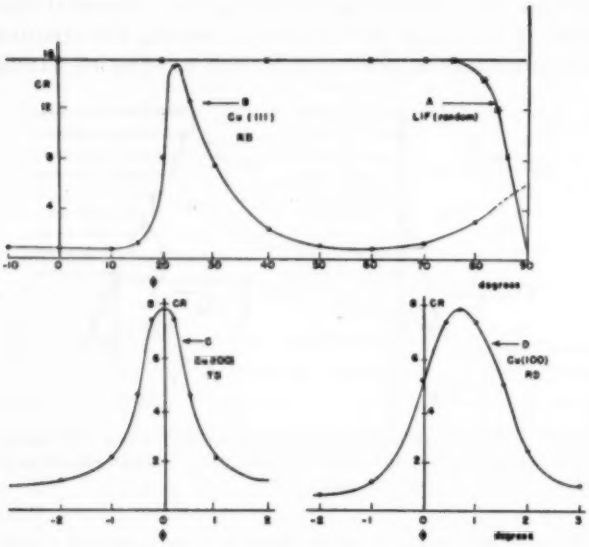


FIG. 4. Experimental results: A, the counting rate, CR, for a sample of powdered LiF using the (111) reflection; B, the result with rolled copper using the (111) reflection and taken in the rolling direction; C, the orientation of recrystallized copper in the transverse direction, and D, the orientation in the rolling direction.

EVALUATION

Figure 4 has shown that the method is accurate and dependable, but it does have several limitations. Most important, in common with all reflection methods, it cannot give information for the part of the pole figure near the edge of the diagram—that is, for values of ϕ in excess of about 80°. To push the method to 85° would require inconveniently narrow slits. Since for some materials it is this edge of the pole figure which is most valuable, the best practical approach is to supplement the reflection data given by this method with that given by transmission. This is a natural thing to do because the transmission method is at its best near the edge of the pole figure. (The transmission method of Asp, Decker, and Harker has been re-examined and the results will be given in a later paper.)

A second limitation of the method in common with all others using a Geiger counter is that the sample volume irradiated is small. In length, the area is about one-fourth inch, and the width equal to the slit width, S_2. Usually the value of S_2 was near 0.020 inches but this width might be doubled if a slight decrease in the constant counting range of ϕ is not objectionable. Such a decrease would be no disadvantage if a combination reflection and transmission procedure is followed.

MATHEMATICAL APPENDIX

In the following explanation for the constant counting rate as the sample is rotated about $F-F'$ in Figs. 1 and 2 the effect of changes in absorption are shown to be exactly canceled by changes in the scattering volume. In Fig. 5 are given two views of the sample surface as seen along $F-F'$. Attention should be directed to the scattering material at the depth, t, in the interval, dt. In A the scattering volume, dV, is equal to $WS_2dt/\sin\theta$ where W is the horizontal width of the incident x-ray beam and S_2 its vertical width. After a rotation as shown in Fig. 5B the volume has increased to a new value,

$$dV=\frac{WS_2dt}{\sin\theta\cos\phi}. \tag{1}$$

The change in absorption can also be found with the help of Fig. 5. In both A and B the incident and diffracted beam make an angle θ with $F-F'$. It is evident, therefore, that the effect of a rotation ϕ is to increase by a factor of $1/\cos\phi$ the path length required to reach a given depth, t, measured normal to the surface.

It is now possible to calculate the reflected intensity, I, from a flat sample placed in the experimental arrangement of Fig. 2. Let I_0 be the incident beam intensity and D the scattering efficiency of the material at the given angle θ. Then,

$$I=I_0D\int[\exp(-2\mu t/\sin\theta\cos\phi)]dV. \tag{2}$$

The integration is to be extended over the entire effective scattering volume of a sample having infinite

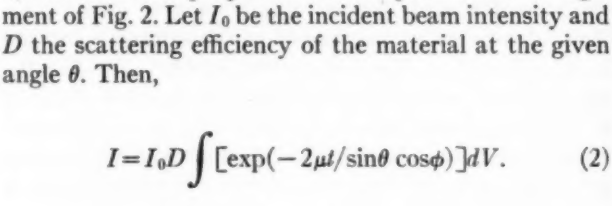


FIG. 5. Diagrams used when calculating the volume and absorption changes during rotation of the sample.

thickness. Using Eq. (1) in Eq. (2),

$$I=\frac{I_0DWS_2}{\sin\theta\cos\phi}\int_0^\infty[\exp(-2\mu t/\sin\theta\cos\phi)]dt. \tag{3}$$

After integrating and simplifying Eq. (3),

$$I=\frac{I_0DWS_2}{2}. \tag{4}$$

This result shows that the intensity of the diffracted beam is independent of ϕ.

For samples having a thickness, T, insufficient to give total absorption, I is not independent of ϕ but is related in the following manner:

$$I=\frac{I_0DWS_2}{2}[1-\exp(-2\mu T/\sin\theta\cos\phi)]. \tag{5}$$

ACKNOWLEDGMENTS

The author wishes to thank Mr. Clifton and Mr. Hess for their assistance during the development of this method and Professor Barrett for suggestions during discussion of the results.

Determination of Preferred Orientation in Flat Transmission Samples Using a Geiger Counter X-Ray Spectrometer

L. G. SCHULZ

Institute for the Study of Metals, University of Chicago, Chicago, Illinois

(Received March 21, 1949)

A method employing a diverging x-ray beam is described for the determination of preferred orientation in flat transmission samples with a Geiger counter x-ray spectrometer. It is shown that for a certain range of sample thickness the counting rate is independent of the rotation of the samples, and therefore, no correction formula is required.

INTRODUCTION

A METHOD for determining preferred orientation in flat transmission samples using a Geiger counter x-ray spectrometer has been described by Asp, Decker, and Harker.[1] The experimental procedure is to first mount the sample in a vertical plane on the axis of the spectrometer. Then, by means of a special mechanism the sample is rotated in its own plane and also about the spectrometer axis. The analysis of this method as given by the authors assumes parallelism of the incident beam, a condition difficult if not impossible to attain in practice. As a result the correction formula derived is not valid except in the range where parallelism is approximately attained. The error arising from using the formula for a non-parallel beam is greatly reduced, however, if a sufficiently wide slit is used at the counter. Although the method is simple and rapid, and within a certain range is also reasonably accurate, it is possible to greatly improve it by making some minor alterations in the slit system. It is the purpose of this paper to apply the general procedure of Asp, Decker, and Harker to the experimental arrangement in which (1) a diverging beam is employed, and (2) a narrow slit is used at the counter entrance. During the discussion the x-ray optics of the Norelco instrument will be assumed.

Experimental results as well as theoretical considerations show that the new arrangement has several advantages. First, it makes the counting rate a symmetrical function of the sample rotation thereby reducing by one-half the amount of calculation. But more important, it was found that for samples within a certain range of thickness the maximum of the counting rate curve could be changed to a broad, flat plateau. In this region no correction formula is required and accordingly the data can be used directly. A special satisfaction resulting from the use of a slit at the counter is that a mathematical analysis can be made of the x-ray optics of the actual physical arrangement rather than of an assumed one. Furthermore, the results of this analysis can be verified with great exactness experimentally. Because of this situation the pole figure data can be secured with a corresponding accuracy.

In an earlier paper on a reflection method for determining preferred orientation it was pointed out that every reflection method is inadequate for obtaining the data required for the edge of a pole figure. Likewise, transmission methods fail to give data for the central region. Since the reflection method is much the easier, the plan suggested is to use it for obtaining most of the data and then employing the transmission method only for completing the edge of the pole figure.

X-RAY SCATTERING FROM A FLAT TRANSMISSION SAMPLE

The experimental arrangement for transmission is shown in Fig. 1. L is the slit source not more than 0.005 in. in width and about 0.050 in. long. From this diverges a wedged-shaped beam whose angular width is controlled by three interchangeable vertical slits. The transmission sample is in the form of a thin sheet mounted with its central plane containing the axis of the spectrometer. With a suitable mechanical arrangement the sample may be turned about this axis by a measur-

[1] Decker, Asp, and Harker, J. App. Phys. **19**, 388–392, (1948).

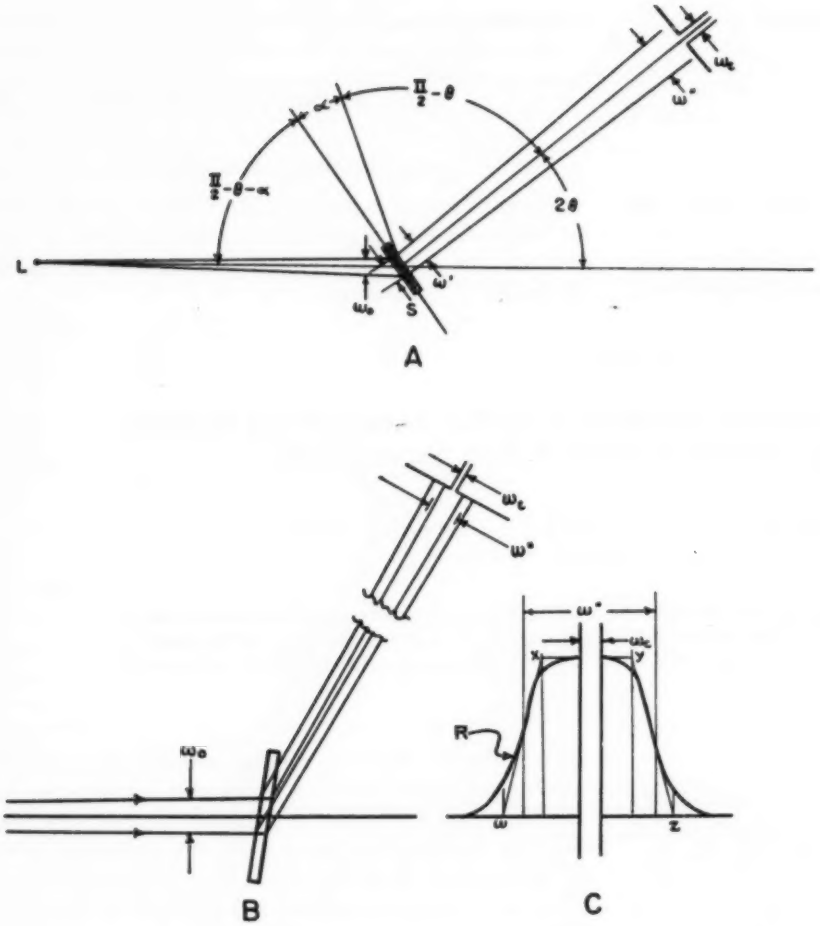


FIG. 1. The plan of the experimental arrangement is shown in A: L is a line source from which diverges a wedge-shaped beam defined by the width w_0. The width of the sample irradiated is S. The width of the diffracted beam at the sample is w' and at the counter it is w''. In B is shown the fringing resulting from the sample thickness. The intensity across the beam at the counter entrance is given in C. Here $wxyz$ is the contour to be expected from geometric considerations while the heavy curve, R, is the contour actually observed experimentally.

able angle α. In addition, the sample can be turned in its own plane. The width of the incident beam as it strikes the sample is w_0, the width of the diffracted beam as it leaves the sample is w', and finally, the width of the diffracted beam when it reaches the counter is w''. The experimental conditions which will be assumed are: (1) the width of the counter slit, w_c, is always smaller than w''; (2) w_c is greater than the width of the source L, and (3) the sample thickness is small relative to w_0. These conditions are natural ones and almost automatically satisfied in practice.

The simple drawing of Fig. 1A implies that the energy

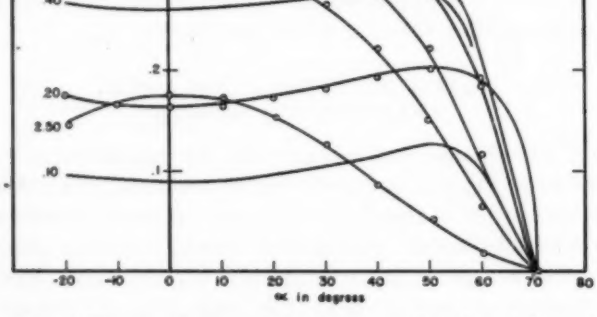


FIG. 2. The counting rate CR is shown as a function of the angle α in Eq. 10. Here θ has been taken equal to 19°. The smooth curves are the calculated values; the circles are the experimental points given by a sample of powdered LiF.

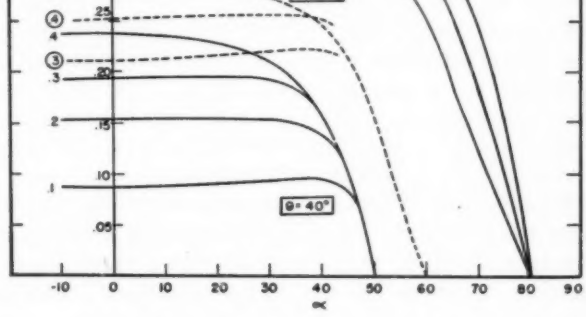


FIG. 3. Counting rate, CR, as a function of α for θ equal to 10°, 30°, and 40°. The object was to find the value of μT for which CR is nearly independent of α.

flux is uniform across the width w'' but in reality there is a fringing at the edges due to the sample thickness. This presents no difficulties, however, because the slit width w_c is kept small enough to avoid the fringing region. In Fig. 1B is shown the beam widths for extreme values of 2θ and α. In practice w_0 would be about 0.050 in. and the sample thickness under 0.005 in. In the drawing, however, a thicker sample is shown to exaggerate the fringing. Because of the divergence of the incident beam it can be seen that at the counter entrance the fringes are relatively narrow even for this thick sample. In Fig. 1C is shown an actual beam contour showing how in a real situation the broadening is greater than that predicted from geometric considerations. The analysis to follow will be valid only over the flat maximum of the curve which corresponds to the central section of the diffracted beam.

The energy entering the counter as a function of the angle will now be calculated. Three factors must be considered: (1) the change in absorption; (2) the change in volume of the scattering material, and (3) the relative widths of w'' and w_c. Since the first two factors have been considered previously, the results will be used directly.[1] If I_0 is the intensity of the incident beam and D the scattering efficiency of the material, then the scattered intensity, I, will be as follows:

$$I=\frac{I_0 D\cos(\theta-\alpha)}{\cos(\theta+\alpha)-\cos(\theta-\alpha)}[\exp\{-\mu T/\cos(\theta+\alpha)\} -\exp\{-\mu T/\cos(\theta-\alpha)\}]. \quad (1)$$

The part of I which will enter the counter is simply the ratio of w_c and w''. w'' can be found from Fig. 1. The irradiated width, S, is equal to $w_0/\cos(\theta+\alpha)$ and the width of the scattered beam, w', is $S\cos(\theta-\alpha)$. Since w'' is equal to the sum of w_0 and w', its value is given by

$$w''=\left[\frac{2\cos\theta\cos\alpha}{\cos(\theta+\alpha)}\right]w_0. \quad (2)$$

The energy, I', entering the counter, which will be $(w_c/w'')I$, can be found from Eqs. (1) and (2). The result given in the next equation is written in a form convenient for calculation and also suitable for pointing out how each of the three factors listed above influence the value of I'.

$$I'=\left[\frac{I_0 D w_c}{w_0}\right]\left[\frac{1}{2\cos\theta\cos\alpha}\right] \times\left[\frac{\exp\{-\mu T/\cos(\theta+\alpha)\}-\exp\{-\mu T/\cos(\theta-\alpha)\}}{\frac{1}{\cos(\theta-\alpha)}-\frac{1}{\cos(\theta+\alpha)}}\right]. \quad (3)$$

In the first term I_0 is proportional to the incident beam width, w_0, consequently I' is independent of w_0, a result which leaves w_c as the effective aperture of the system. The second term, $1/2\cos\theta\cos\alpha$, is equal to S/w'', therefore, it contains the combined effect of changes in scattering volume and changes in the beam width, w''. This term is symmetrical about α equal to zero as is also the last term which arises because of absorption changes. For very small values of μT this absorption term reduces to μT and accordingly for thin samples I' is proportional to $Tw_c/\cos\alpha$. In the experimental results to be shown I' was not measured absolutely but as a linear function of the counting rate, CR.

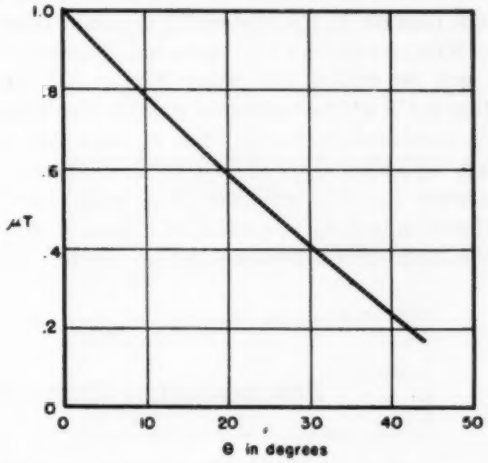


FIG. 4. A graph showing the optimum value of μT for various values of θ. With these values of μT, CR is nearly independent of values of α up to about 25°.

EXPERIMENTAL VERIFICATION AND PRACTICAL CONSIDERATIONS

Powder samples having random orientation were prepared by pressing LiF powder into suitably shaped holes through sheet metal of various thicknesses. These samples were mounted with their median plane containing the axis of the spectrometer. In Fig. 2 is given a typical set of results in which CR is shown as a function of α. The continuous curves are the calculated values; the circles are the values obtained experimentally. The agreement is seen to be good over a great range of sample thickness.

The curves of Fig. 2 suggest a possible method of avoiding the use of a correction formula. When μT is less than 1.0 all the curves show a shallow minimum at α equal to zero. Accordingly, a sample having a random orientation could be rotated about the vertical axis without changing appreciably the counting rate CR. A further inspection of the curves shows that for μT near 0.6 the amplitude of α can be as much as 45° without changing CR by more than a few percent. In Fig. 3 similar curves are shown for other values of θ. It is evident that for each θ a μT can be selected to give a constant CR.

Consideration of the curves of Fig. 2 and Fig. 3 has led to these practical conclusions.

Procedure A.—If the transmission method is used to supplement the reflection method which was described earlier, the amplitude of α need not exceed 20°. The curves show that if μT is between 0.3 and 0.4 the variation in CR does not exceed 5 percent for any value of θ up to 40°. With one sample having this required thickness it is possible to use any of the common reflections (111), (200), or (220) for f.c.c. material. Normally the θ-value will lie within the range 20° to 40° and the variation in CR will be somewhat smaller than 5 percent. This is considerably better than is required, or can normally be achieved by photographic methods.

Procedure B.—If one reflection is being used—one value for θ—and if it is desirable to extend α up to 40°, then the optimum value for μT can be selected from the CR curves. In Fig. 4 a graph is given which shows how μT varies with θ. With this procedure one sample can be used for the (111) and (200) reflections but a second, a thinner, sample is required for the (220).

Procedure C.—If it is inconvenient to obtain a sample in the range 0.3 to 0.4 then the entire CR curve can be plotted and the experimental data corrected before being used in plotting a pole figure.

ACKNOWLEDGMENT

The author wishes to thank Mr. Clifton and Mr. Hess for assistance during the development of this method, and Professor Barrett for suggestions during the discussion of the results.

Propagation of Transverse Waves in Viscoelastic Media*

F. T. Adler, W. M. Sawyer,** and John D. Ferry

Departments of Physics and Chemistry, University of Wisconsin, Madison, Wisconsin

(Received February 17, 1949)

In calculating elastic constants of viscoelastic media from transverse wave propagation, it has been customary heretofore to assume a plane disturbance in a medium of infinite extent. Equations have now been derived for conditions which approach much more closely the experimental arrangement, namely, propagation from a finite source plate in a rectangular cell, and representative numerical calculations have been made. The new theory shows that the finite dimensions of cell and source should have very little influence on the wave-length as determined from observations of strain double refraction, but the attenuation should be perceptibly different from that of the plane wave case. Experimental data confirm these conclusions. Nevertheless, the simple plane theory is adequate within present experimental error for the majority of cases. Only in certain specified ranges of mechanical consistency will it be necessary to apply corrections taking into account the sizes of the cell and source.

INTRODUCTION

MEASUREMENTS of wave-length and attenuation of transverse waves in viscoelastic media are useful for calculating fundamental elastic constants of such systems. From the dependence of the elastic constants on frequency, temperature, and other variables, conclusions may be drawn concerning the internal structure of the media studied.

The method[1] has been applied to a variety of concentrated polymer solutions[2] and gels.[3] The medium (ordinarily incapable of supporting its own weight) is contained in a rectangular cell, and waves are set up by a thin plate undergoing sinusoidal oscillations in its own plane (Fig. 1). Here the oscillation is in the z direction, and the waves are propagated in the x direction. Observation of the waves depends upon the strain double refraction produced. When the cell, together with a double quartz wedge, is placed between crossed Polaroid plates and illuminated by a stroboscope synchronized with the vibration frequency, a stationary pattern is observed (Fig. 2); the vertical displacement in each wavy line is a function of the relative retardation (double refraction) introduced in passing through the solution.[1] If the double refraction is proportional to the strain $(\partial u/\partial x)$, it is then possible to calculate from experimental observations a quantity (δ) which is proportional to the integral with respect to y of the component of shear strain perpendicular to the y direction, i.e.,

$$\delta \propto \int (\partial u/\partial x) dy, \qquad (1)$$

where u is the displacement in the medium. The mechanical constants of the medium can then be determined from an expression for u as a function of the space coordinates and the parameters characteristic of wave motion.

* Supported in part by the Research Committee of the Graduate School of the University of Wisconsin from funds supplied by the Wisconsin Alumni Research Foundation.

** Present address: Shell Development Company, Emeryville, California.

[1] J. D. Ferry, Rev. Sci. Inst. **12**, 79 (1941).

[2] J. D. Ferry, J. Am. Chem. Soc. **64**, 1323 (1942); J. N. Ashworth and J. D. Ferry, *ibid.* **71**, 622 (1949); unpublished work by W. M. Sawyer.

[3] J. D. Ferry, J. Am. Chem. Soc. **70**, 2244 (1948); J. D. Ferry and P. R. Morrison, *ibid.* **69**, 388 (1947); J. D. Ferry and J. E. Eldridge, J. Phys. Coll. Chem. **53**, 184 (1949).

In previous applications of this method, it has been assumed that the disturbance could be regarded as a plane wave in a medium of infinite extent. In this case δ is given by

$$\delta=\delta_0 \exp\{i(\omega t-2\pi x/\lambda)-x/x_0\} \tag{2}$$

and the constants λ and x_0 can be readily evaluated from experimental data. (Because of the synchronization, the term ωt merely introduces an arbitrary phase shift.) In this case, u and $\partial u/\partial x$ are functions of x only; δ is proportional to $\partial u/\partial x$, and the observed values of λ and x_0 can be taken as the wave-length and critical damping distance for the shear wave itself. For such a plane wave, the mechanical constants of the medium are readily calculated[4] from λ and x_0 (see below).

SOLUTION OF THE WAVE EQUATION FOR TWO DIMENSIONS

The experimental conditions do not, of course, correspond to a medium of infinite extent; it is the purpose of this paper to discuss the behavior of δ and the modifications in the plane wave solution arising from the finite dimensions of the cell and vibrating plate.

The wave equation for the displacement can be written as

$$\nabla^2 u+\beta^2 u=0, \tag{3}$$

where

$$\beta\equiv(\beta_1-i\beta_2)=\left\{\frac{\rho\omega^2}{G'+i\omega\eta'}\right\}^{\frac{1}{2}}. \tag{3-1}$$

Here, G' is the component of stress in phase with the strain divided by the strain, and η' is the component of stress in phase with the rate of strain divided by the rate of strain; ρ is the density of the medium.

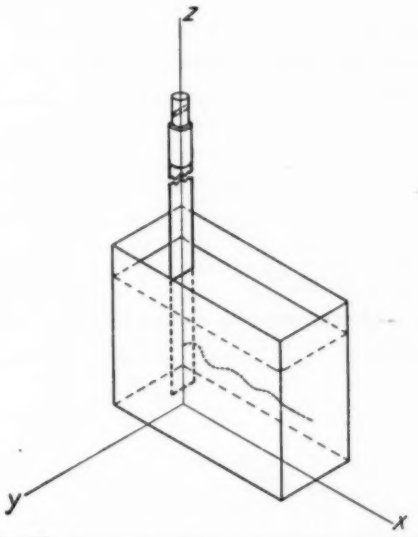


FIG. 1. Isometric drawing of experimental arrangement for wave propagation in a rectangular cell.

[4] Ferry, Sawyer, and Ashworth, J. Polymer Sci. 2, 593 (1947).

The boundary conditions for (3) require vanishing displacement at the walls of the cell and a displacement u_0 at the vibrating plate. Let $2y_0$ and $2y_1$ be the widths of the plate and cell, respectively; then we have

$$u(x,\pm y_1)=0, \tag{4-1}$$

$$u(x,y)\to 0, \text{ for large } x, \tag{4-2}$$

$$u(0,y)=\begin{cases} u_0 \text{ for } |y|\leqslant y_0 \\ 0 \text{ for } |y|>y_0 . \end{cases} \tag{4-3}$$

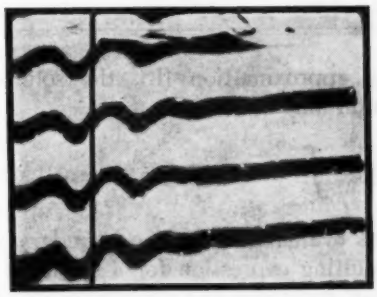


FIG. 2. Pattern of strain double refraction obtained by stroboscopic illumination (17 percent polyvinyl acetate in 1,2,3-trichloropropane, 21.1°C, frequency 400 cycles/sec.).

In general, the substances investigated have sufficiently high damping so that within experimental accuracy the effect of the far wall parallel to the yz plane can be neglected; and we may consider the cell as infinitely extended in the x direction. Equations (4–1) through (4–3) represent an idealization, since it is assumed that the medium in the plane of the plate, outside the plate, is not affected by the motion of the latter. While this idealization will certainly not be valid in the immediate neighborhood of the plate, the effects will die out in the region of the medium where measurements are taken. This assumption will be discussed further.

Within the experimental accuracy no effects arising from the finite height of the medium are observed, since patterns taken at different heights in the cell are sensibly identical. This was confirmed by obtaining a solution for the three-dimensional problem and making a few computations which indicated the dependence on z to be small. We shall therefore consider only the two-dimensional case, $u=u(x,y)$.

The following approximations to the solution of (3) will be considered:

(a) Plane wave solution: Considering cell and plate as infinite, we obtain (2); this is the solution used at present for discussion of experimental data.[2,3]

(b) Finite source plate of infinite height in an infinite medium: Neglecting the lateral boundaries, the contributions to δ coming directly from the plate are obtained, while the contributions of the reflections at the cell walls are neglected; it is assumed that u vanishes in the plane of the plate, except on the plate.

(c) Finite source in a finite medium: Considering only the effects of the lateral walls we use the boundary conditions (4–1) through (4–3).

TABLE I. Values of $\beta_1 x/\pi$ at which $\delta=0$, calculated from Eq. (7). $y_0=0.4$ cm., $y_1=1.25$ cm.

β_1, cm^{-1}	$\beta_2/\beta_1=0.05$	$\beta_2/\beta_1=0.10$	$\beta_2/\beta_1=0.20$	$\beta_2/\beta_1=0.30$
5		2.00	1.97	1.93
		3.06	3.03	3.00
		4.12	4.09	4.04
		5.16	5.11	5.06
		6.18	6.15	6.09
10	1.98	1.97	1.94	1.91
	2.98	2.96	2.94	2.91
	3.97	3.96	3.94	3.92
	4.96	4.95	4.95	4.92
	5.96	5.95	5.95	5.93

Assuming approximation (b), the solution of the wave equation may be written

$$u(x,y)=\int_{-\infty}^{+\infty} dt\, a(t)\exp[-x(t^2-\beta^2)^{\frac{1}{2}}]\cos ty,$$

where $a(t)$ is evaluated from the boundary conditions (4). The resulting expression for δ in closed form, involving Bessel functions and complex error integrals, is cumbersome. The problem of finite cell size may be approximated by an image method, and the solution shows that for large x the amplitude of δ decreases in the region of experimental interest more slowly than an exponential; thus the slope of the envelope of δ may be expected to differ in general from that obtained from (a). In view of the results obtained from (c) and of their interpretation, it does not seem necessary to give details of this solution at present.

Assuming approximation (c), the solution of the wave equation was obtained in the form of a Fourier series satisfying the boundary conditions (4–1) to (4–3):

$$u(x,y)=\frac{4}{\pi}u_0\sum_{n=0}^{\infty}\frac{1}{(2n+1)}\exp(-\alpha_n x)\sin(n+\tfrac{1}{2})\frac{\pi}{y_1}y_0 \times\cos(n+\tfrac{1}{2})\frac{\pi}{y_1}y, \quad (5)$$

where

$$\alpha_n=[-\beta^2+(n+\tfrac{1}{2})^2(\pi^2/y_1^2)]^{\frac{1}{2}}. \quad (6)$$

From (1) and (5) we have

$$\delta=\frac{16}{\pi^2}u_0 y_1\sum_{n=0}^{\infty}c_n\exp(-\alpha_n x), \quad (7)$$

where

$$c_n=\frac{(-1)^{n+1}}{(2n+1)^2}\alpha_n\sin(n+\tfrac{1}{2})\frac{\pi}{y_1}y_0. \quad (8)$$

Results of a numerical evaluation of (7) will be discussed in the next section.

The influence of the approximation implied in the boundary condition (4.5) was considered, and it can be shown that at a moderate distance from the discontinuity, its influence on the computed value of δ will not be appreciable,[5] i.e., calculations based on (4.5) and on the assumption of a smooth decay in the influence of the plate in its own plane in the region $|y|>y_0$ will yield substantially the same value for δ.

DISCUSSION OF NUMERICAL RESULTS

The parameters β_1 and β_2 characterize the viscoelastic properties of the medium. The relation (3–1) implicitly defines the constants G' and η' by

$$G'\equiv\rho\omega^2\frac{\beta_1^2-\beta_2^2}{[\beta_1^2+\beta_2^2]^2}, \quad (9\text{–}1)$$

$$\eta'\equiv 2\rho\omega\frac{\beta_1\beta_2}{[\beta_1^2+\beta_2^2]^2}. \quad (9\text{–}2)$$

In terms of the plane wave approximation parameters, the quantities β_1 and β_2 have a simple significance:

$$\{2\pi i/\lambda+1/x_0\}^2=-\beta^2=(i\beta_1+\beta_2)^2.$$

Thus $\beta_1=2\pi/\lambda$ and $\beta_2=1/x_0$; substitution of these symbols into (9) gives G' and η' in terms of λ and x_0.

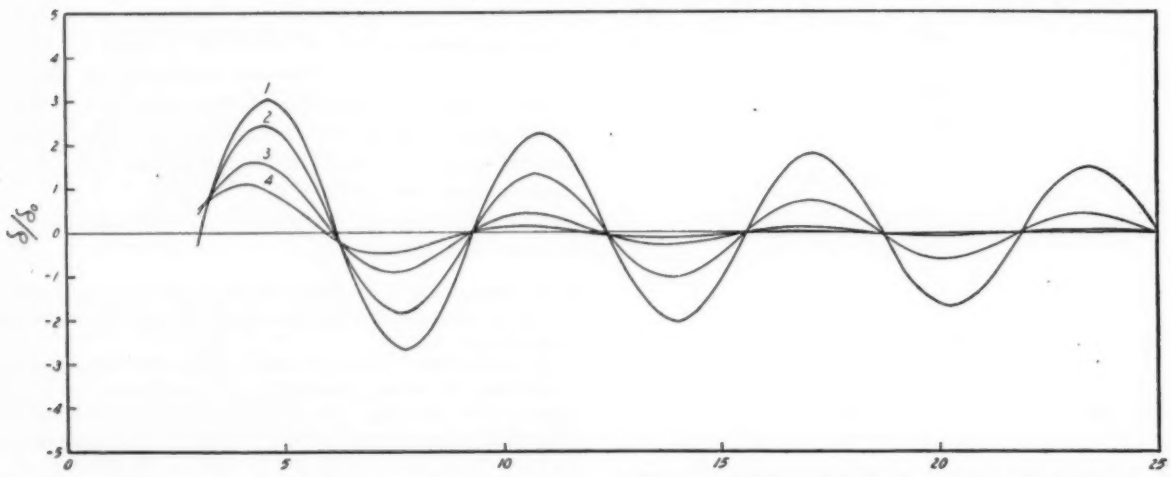


FIG. 3. Plot of δ/δ_0 against $\beta_1 x$, calculated from Eq. (7). $y_0=0.4$, $y_1=1.25$, $\beta_1=10$. Values of β_2/β_1: 1, 0.05; 2, 0.10; 3, 0.20; 4, 0.30.

[5] Compare J. D. Tamarkin and W. Feller, "Partial differential equations," Brown University Notes, p. 84 ff.

In practice, the experimental data are fitted by forcing them into this approximation, i.e., with an expression of the form

$$\delta = \delta_0 \exp(2\pi i x/\lambda^* - x/x_0^*), \tag{10}$$

where the starred parameters are obtained by the forced fit. It is true that for finite cells and plates δ is given by a sum of exponentials and it is obvious from (7) that (10) cannot hold exactly. However, it will be shown that for most cases the direct identification of x_0^* and λ^* with x_0 and λ will yield results for G' and η' within the limits of experimental error as specified in the section on experimental data below.

To investigate the accuracy of the approximation (10), numerical calculations of δ were made from (7), for representative values of y_0, y_1, β_1, and β_2 corresponding to the dimensions of our present equipment and the range of mechanical consistencies currently studied. Examples of δ plotted against the dimensionless variable $\beta_1 x$ are shown in Fig. 3 for several values of β_2, the other parameters being constant. From plots of this type the zeros of δ were determined. The nodes were evenly spaced; the deviation of the difference of the values of $\beta_1 x/\pi$ for successive zeros of δ from unity was of the order of 0.02 or less provided that the wave-length was not too large, i.e., $\beta_1 > 5$, or that the damping was not negligible, i.e., $\beta_2/\beta_1 > 0.1$; here we exclude the region next to the plate where edge effects are important and where the data are not reliable, and consider only the region from the first node outward. Representative values of the zeros are given in Table I. It is thus justifiable, within the range covered here, to treat $2\pi/\beta_1$ as the wave-length of the strain wave in the medium, and thus to identify the experimental λ^* with $2\pi/\beta_1$; i.e., λ^* is substantially equal to λ.

We may therefore expect that δ obtained from (7) will be approximated very closely by

$$\delta = A(x)e^{2\pi i x/\lambda}. \tag{11}$$

Next, the envelope of δ, derived from (7), was plotted logarithmically against $\beta_1 x$ to obtain the function $A(x)$. Figure 4 shows a representative example of these curves. Although log$A$ *versus* $\beta_1 x$ may usually be fairly well fitted by a straight line, as approximation (10) demands, it is obvious from Fig. 4 that the slope is not constant between $\beta_1 x = 10$ and $\beta_1 x = 20$; that is, the introduction of x_0^* in (10) is a less adequate approximation than the introduction of λ^*.

To gauge the departure of (7) from (10), as measured by the curvature of plots like Fig. 4, the slope of the secant between the i^{th} and j^{th} maximum of δ, denoted by γ_{ij}, was computed for three pairs of maxima. Data of the type of Table II show that γ_{17} and γ_{67} differ by less than 5 percent if the damping is substantial, i.e., $\beta_2/\beta_1 \geq 0.15$; they are also essentially equal for very short wave-length, corresponding to $\beta_1 \sim 15$. Under these conditions the forced fit of $A(x)$ to the exponential function $A_0 e^{-x/x_0^*}$, corresponding to (10), and the identification of x_0^* with $1/\beta_2$, is satisfactory within experimental error. At a great distance from the plate, the solution (7) will reduce to a single exponential, the higher harmonics having been damped out. The amplitude of δ at this distance will decrease like $e^{-\beta_2 x}$, and the slope of logδ is then actually equal to β_2. However, the sample data show that particular care has to be taken in the case of long wave-lengths, because in this case the complete damping out of the

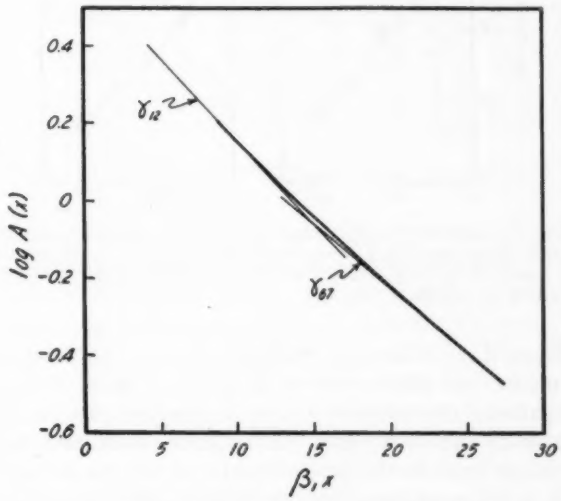


FIG. 4. Plot of log$A(x)$ against $\beta_1 x$, calculated from Eq. (7). $y_0 = 0.4$, $y_1 = 1.25$, $\beta_1 = 10$, $\beta_2/\beta_1 = 0.10$.

TABLE II. Slopes of log$A$ plotted against $\beta_1 x$. $y_0 = 0.4$ cm., $y_1 = 1.25$ cm.

			$\beta_2/\beta_1 = 0.05$		$\beta_2/\beta_1 = 0.10$		$\beta_2/\beta_1 = 0.20$		$\beta_2/\beta_1 = 0.30$	
β_1, cm^{-1}	i	j	γ_{ij}	$(\beta_2/\beta_1)/\gamma_{ij}$	γ_{ij}	$(\beta_2/\beta_1)/\gamma_{ij}$	γ_{ij}	$(\beta_2/\beta_1)/\gamma_{ij}$	γ_{ij}	$(\beta_2/\beta_1)/\gamma_{ij}$
5	1	2			0.0732	1.37	0.183	1.09	0.286	1.05
	6	7			0.0908	1.10	0.205	0.98	0.312	0.96
	1	7			0.1031	0.97	0.206	0.97	0.307	0.98
10	1	2	0.0517	0.97	0.100	1.00	0.199	1.00	0.303	0.99
	6	7	0.0299	1.67	0.0787	1.27	0.188	1.06	0.295	1.02
	1	7	0.0380	1.32	0.0892	1.12	0.191	1.05	0.293	1.02
20	1	2			0.101	0.99	0.203	0.99	0.303	0.99
	6	7			0.103	0.97	0.203	0.99	0.294	1.02
	1	7			0.100	1.00	0.200	1.00	0.299	1.00

higher harmonics will not take place in a cell of the dimensions used currently. For increased accuracy, if the experimental precision should warrant it, the slope γ_{ij} (of some localized secant) could be measured and multiplied by the calculated ratio $(\beta_2/\beta_1)/\gamma_{ij}$ (see Table II) as a correction factor to give β_2/β_1.

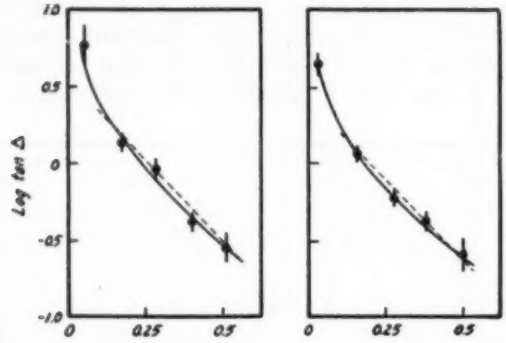


FIG. 5. Experimental damping curves for 17 percent polyvinyl acetate (Fraction I) in 1,2,3-trichloropropane, 25.6°C, 400 cycles/sec. $y_1=1.25$. Left: $y_0=0.4$, $\lambda^*=0.59$, $x_0^*=0.54$. Right: $y_0=1.25$, $\lambda^*=0.59$, $x_0^*=0.51$.

Even if $A(x)$ deviates markedly from an exponential form, so that plots of the type of Fig. 4 show marked curvature, the periodicity of δ remains unchanged (Table I) and so λ^* may still be safely identified with $2\pi/\beta_1$, at least in the case of media of current interest.

It is thus concluded that at the present level of precision and in media of current interest the plane approximation gives very satisfactory values of β_1, while β_2 is usually within experimental error but must be regarded with more caution. The mechanical constants sought, G' and η', depend on both β_1 and β_2. However, G' is never in serious doubt. For $\beta_2/\beta_1 \geqslant 0.15$ (high damping) both β_2 and β_1 are reliably obtained from (10); and for $\beta_2/\beta_1 < 0.15$ (low damping) Eq. (9-1) may be expanded in β_2/β_1 to give $G' \simeq (\rho\omega^2/\beta_1^2)\{1-3(\beta_2/\beta_1)^2 + \cdots\}$, showing that an inaccuracy in β_2 will only enter in a term contributing less than 7 percent to G'. On the other hand, η' is proportional to β_2 and reflects in full any uncertainty in the latter.

COMPARISON WITH EXPERIMENTAL RESULTS

The conclusion that λ^*, measured from experimental data, can be identified with $2\pi/\beta_1$ is amply justified by experience. The observed wave-lengths are always independent of the distance from the vibrating source within experimental error (ordinarily ± 5 percent). In a number of cases comparisons of λ^* have been made for the same medium in different cells or with vibrator plates of different sizes (Tables III and IV); the agreement is nearly always within 5 percent and often much better.

The conclusion that x_0^*, measured from a forced linear fit to a plot of $\log A(x)$, can be approximately identified with $1/\beta_2$ but that the fit may be somewhat ill-defined is also borne out by experiment. Typical experimental damping curves, derived from the envelope of the observed strain wave, are shown in Figs. 5, 6, and 7. Here the ordinate, log tanΔ, differs from $A(x)$ by an additive constant.[1,2] The vertical line through each point represents the error. The first curve in Fig. 5, and that in Fig. 7, correspond nearly to the conditions of the theoretical curve of Fig. 4 ($y_0=0.4$, $y_1=1.25$, $\beta_1=10$, $\beta_2/\beta_1=0.1$). In Fig. 5, $\beta_1=11$ and $\beta_2/\beta_1=0.2$; in Fig. 7, $\beta_1=9.2$ and $\beta_2/\beta_1=0.09$. Their shapes are indeed very similar, as are those of the other experimental curves, exhibiting slight concavity upward.

To obtain x_0^*, the best straight line is drawn through the points, as shown by the dotted lines in the figures. The first point is always omitted, since the calculations have shown that its deviation is the greatest (γ_{12} often departs markedly from β_2/β_1) and moreover its measurement is subject to more uncertainty. The values from the two curves in Fig. 5, corresponding to different sizes of source plates, agree within 6 percent. Tables III and IV give a more extended series of comparisons of x_0^* values obtained with different cell and plate dimensions. They agree within ± 15 percent, and often better; as predicted, not nearly so well as λ^*, but still within the rather large experimental error. It appears not yet worth while here to apply the next approxima-

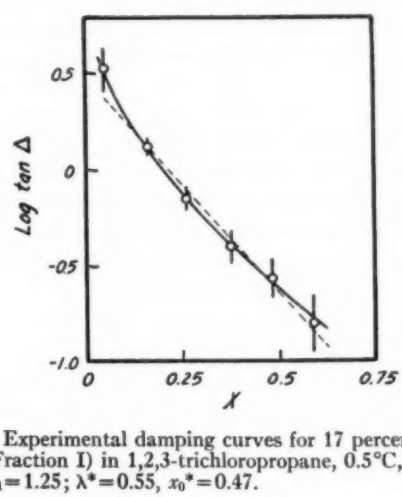


FIG. 6. Experimental damping curves for 17 percent polyvinyl acetate (Fraction I) in 1,2,3-trichloropropane, 0.5°C, 400 cycles/sec. $y_0=y_1=1.25$; $\lambda^*=0.55$, $x_0^*=0.47$.

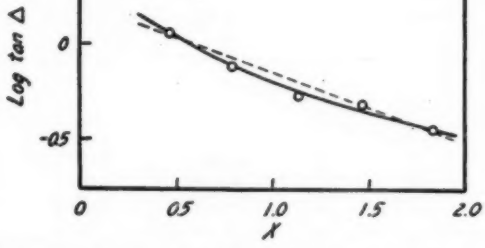


FIG. 7. Experimental damping curves for 25 percent polyisobutylene in xylene, 28.0°C, 500 cycles/sec. $y_0=0.4$, $y_1=1.25$, $\lambda^*=0.68$, $x_0^*=1.17$.

TABLE III. Comparisons of λ^* and x_0^* observed in different cells with the same plate. $y_0=0.4$; a: $y_1=0.5$; c: $y_1=1.25$.

Polymer-solvent	Conc., wt. %	Temp. °C	ν cycles/sec.	$\lambda^*(a)$	$\lambda^*(c)$	$x_0^*(a)$	$x_0^*(c)$
Polyisobutylene-xylene[a]	15.3	28.0	160	0.87	0.92	—	—
			250	0.67	0.67	0.70	0.79
			320	0.52	0.51	0.44	0.58
			400	0.43	0.40	0.38	0.48
			500	0.35	0.35	0.35	0.42
Polyvinyl acetate-1,2,3-trichloro-propane[b]	35.1	0.0	1000	0.73	0.69	0.49	0.47
			1600	0.48	0.44	0.32	0.29
			2000	0.38	0.35	0.24	0.20
		31.0	800	0.65	0.62	0.49	0.50
			1000	0.54	0.51	0.41	0.39
			1250	0.43	0.41	0.27	0.26

[a] Data of J. N. Ashworth.
[b] Data of W. M. Sawyer.

TABLE IV. Comparison of λ^* and x_0^* observed with different plates in the same cell. $y_1=1.25$; c: $y_0=0.4$; d: $y_0=1.25$.

Polymer-solvent	Conc., wt. %	Temp. °C	ν cycles/sec.	$\lambda^*(c)$ cm.	$\lambda^*(d)$ cm.	$x_0^*(c)$ cm.	$x_0^*(d)$ cm.
Polyisobutylene-xylene[a]	18.7	25.0°	160	1.35	1.37	—	—
			200	1.10	1.12	—	—
			250	0.84	0.86	0.93	0.86
			320	0.68	0.70	0.81	0.73
			400	0.55	0.56	0.73	1.00
			500	0.47	0.47	0.73	0.66
Polyvinyl acetate-1,2,3-trichloro-propane[b]	17.0	25.6	160	1.40	1.40	0.87	0.86
			200	1.04	1.09	0.66	0.69
			250	0.92	0.89	0.68	0.67
			320	0.73	0.73	0.61	0.63
			400	0.59	0.59	0.51	0.54
			500	0.47	0.48	0.43	0.38
			630	0.38	0.40	0.32	0.29
			800	0.31	0.32	0.23	0.22
			1000	0.25	0.28	0.20	0.20

[a] Data of J. N. Ashworth.
[b] Data of W. M. Sawyer.

tion of estimating γ_{67}, or another specific secant, and using a correction factor.

CONCLUSIONS

Numerical values for the distance between nodes and for the damping of a shear wave progressing in a finite cell have been obtained for representative values of G' and η', characterizing viscoelastic media of current interest. These solutions of the wave equations show that the current method of evaluating data by applying a forced-fit solution of the plane wave type will lead to values of G' which are reliable within the experimental limits of error. As it is shown by the calculations, and borne out by experiment, the first maximum of the shear wave should be omitted in evaluating the damping, because it is normally too high. The value of η' thus obtained will reflect in full the error introduced by the forced-fit solution, but for media of current interest it will not exceed 15 percent. The forced fit will always give a good approximation in the case of moderate or high damping. For low damping, an additional error, not considered theoretically, is introduced by the fact that reflections from the far end of the cell are neglected, and that the wave will become exponentially damped only after having traveled many wave-lengths in the cell. Thus, the forced-fit method may become unreliable for long wave-lengths, and, as a result, data obtained in cells of different sizes may not agree.

It should be emphasized that with increasing precision of the measurements, or change to media with a new range of mechanical constants, the plane wave approximation is very likely to become inadequate for the evaluation of data; in this case, a more elaborate numerical analysis may become necessary.

ACKNOWLEDGMENTS

The authors would like to express their appreciation of the support given to this work by the Wisconsin Alumni Research Foundation, and to the Computing Service of the University of Wisconsin for its cooperation. They are indebted in particular to Professor K. J. Arnold and Mrs. Marie H. Glissendorf for performing the numerical calculations and preparing the tables contained in this note. They also want to thank Dr. J. N. Ashworth for permission to quote his experimental data.

A Study of the Mechanism of Sintering of Metallic Particles

J. H. Dedrick* and A. Gerds**
University of Cincinnati, Cincinnati, Ohio
(Received March 23, 1949)

An investigation has been carried out on the mechanism of sintering of metallic particles. Single layers of spherical copper particles have been sintered under varying conditions of time and temperature in a dry hydrogen atmosphere. Analysis of the rate of interface contact between the particles has shown metallic sintering to be a diffusion process, in agreement with the theory of Kuczynski. The heat of activation of self-diffusion of copper has been calculated and found to be 55,000 cal./mole. This is in close agreement with Kuczynski and also values obtained by other workers using the radioactive tracer method.

INTRODUCTION

THE mechanism of sintering of metallic powders is a phenomenon on which complete agreement has not been attained as yet. Numerous attempts have been made to develop a theory of sintering, J. Frenkel[1] being one of the first to work along these lines. He concluded that the sintering process was one in which viscous flow of crystalline bodies took place under the influence of surface tension. Frenkel derived an equation relating the growth of the interface between two spherical crystalline particles or between a particle and a semi-infinite crystal to the time t at constant temperature. The relationship can be written as follows:

$$x^2=3/2(a\sigma/\eta)t \qquad (1)$$

where x is the radius of the interface assumed to be circular, a the original radius of the sphere, σ the surface-tension of the material and η the viscosity. The other assumptions are that x/a is less than 0.3 and that during the period t, the original radius, a, of the metallic particle did not change.

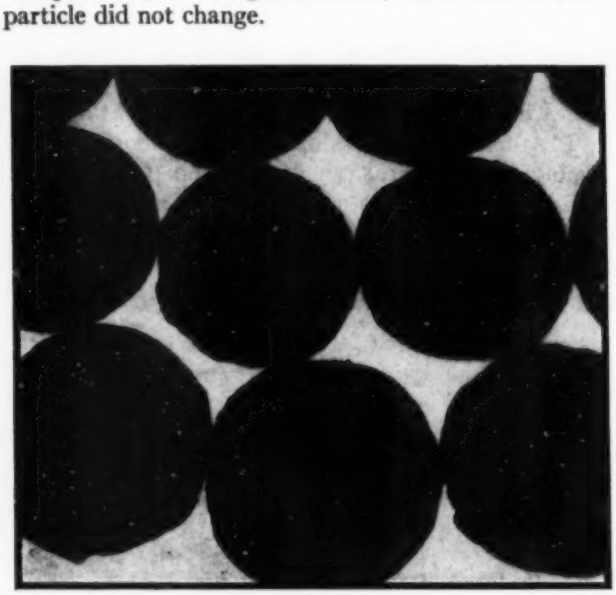


Fig. 1. Copper particles sintered in dry hydrogen at 700°C for 50 hours (46×).

A. J. Shaler and J. Wulff,[2] following Frenkel's viscous flow line of thought, developed a comprehensive theory of sintering of a mass of metallic powder, this being the first attempt to put mass sintering on a scientific basis. G. Kuczynski[3] developed a theory of sintering of two spherical crystalline particles, based on volume diffusion, which resulted in the following relationship:

$$x^5/a^2=(40\sigma\delta^3/kT)D_v t \qquad (2)$$

where x is the radius of the interface, assumed to be circular, a the original radius of the sphere, σ the surface tension of the material, δ the interatomic distance, k the Boltzmann constant, T the absolute temperature, t the time, and D_v the coefficient of volume self-diffusion. This relationship shows that if volume diffusion is the mechanism of sintering, the fifth power of the radius of the interface between the particles should be proportional to the time of sintering.

Very little conclusive experimentation was carried out except for the extensive work of Kuczynski. By sintering spherical copper powder to a copper block and silver powder to a silver block at various temperatures, in hydrogen, for different periods of time, he was able to experimentally verify his diffusion theory. Plotting $\log(x/a)$ *versus* $\log t$ for different temperatures, he obtained straight lines, the slope being approximately 5 in all cases. Plotting $\ln D_v$ against $1/T$ for copper, he obtained 56,000 calories/mole for Q_v the heat of activation of self-diffusion. By similar experimentation, he obtained 42,000 calories/mole for Q_v for silver. These values are well in accord with those already obtained by radioactive tracer methods.

The present investigation was carried out at the University of Cincinnati, on copper powder, using different experimental techniques than heretofore used, with the aim in mind being to obtain further information on the sintering of crystalline bodies, and also to lend substantiation or repudiation to the theories already proposed.

EXPERIMENTAL METHOD

In this investigation atomized spherical copper powder was used. The powder was obtained from the

* Now at Sylvania Electric Products, Inc., Bayside, New York.
** Now at Battelle Memorial Institute, Columbus, Ohio.
[1] J. Frenkel, J. Phys. U.S.S.R. 9, No. 5, 385 (1945).
[2] A. J. Shaler and J. Wulff, Ind. Eng. Chem. 40, 838 (1948).
[3] G. C. Kuczynski, J. Metals (Feb. 1949), p. 169.

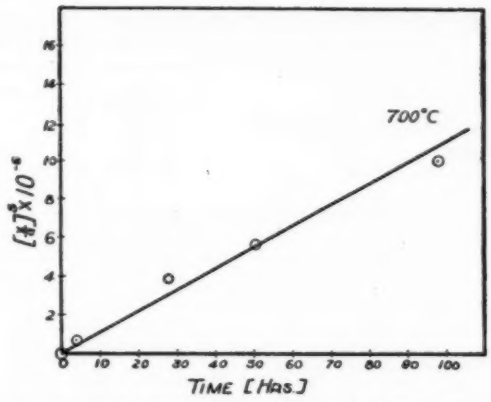


FIG. 2. Plot of the fifth power of the ratio of the average interface radius $\bar{x}$ to the average particle radius $\bar{a}$ against time. Sintering temperature, 700°C.

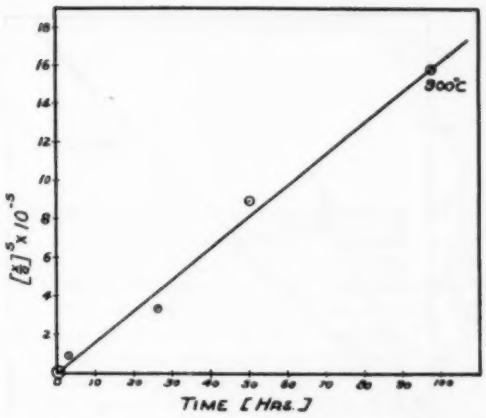


FIG. 3. Plot of the fifth power of the ratio of the average interface radius $\bar{x}$ to the average particle radius $\bar{a}$ against time. Sintering temperature, 900°C.

Metals Disintegrating Company, and the particles selectively picked for the experimental work varied from about 560 to 640 microns in diameter.

The sintering technique used consisted of assembling a close packed single layer of particles on a graphite block and sintering them in dry hydrogen at different temperatures for varying periods of time. Prior to sintering in hydrogen the particles were separated and deoxidized in a carbonaceous atmosphere at 900°C for 24 hours. Figure 1 shows a partial section (magnified 46 times) of a single layer of particles after sintering at 700°C for 50 hours. The sphericalness of the particles used can be ascertained from this picture.

The holder for the copper particles consisted of a small graphite block with a rectangular shaped recess in the top in which the particles were placed. This recess had a very smooth surface on the bottom to prevent particles from lodging in any crevices and as a secondary precaution the bottom of the recess was tapered slightly from the horizontal, in order to keep the individual particles in good contact with one another.

After heating a layer of particles at a given temperature, for a certain period of time, the layer was carefully removed from the graphite block and photographed at 46×, care being taken to focus at the interface between the particles. Then the layer was put back on the graphite block, and sintered at the same temperature for an additional amount of time. This procedure was carried out on copper particle layers sintered at 700°C, 900°C, 1000°C, and 1025°C. The times of sintering, for any given temperature, at which measurements were taken were approximately 3, 25, 50 and 95 hours, except for the 1025°C run, for which measurements were taken at 3 hours and 25 hours.

Measurements were taken of the particle radii and the interface radii at 176×. This was accomplished by projection enlargement of the 46× negatives. Average values, so obtained, of the interface radii, x, and the particle radii a are given in Table I. Each value for $\bar{a}$ is the average of about thirty measurements. For a given time at a particular sintering temperature, the value of $\bar{x}$ was arrived at by averaging the minimum interface radii, the number averaged being about twenty. This procedure was resorted to since the particles were not perfect spheres, and as a result had certain flat spots on their surfaces. During packing of the spheres, prior to sintering, it was impossible to completely prevent two flat spots on adjacent spheres from coming into contact, which of course would result in the obtaining of extremely high interface radius values should such contacts be measured and averaged during the experimentation. By taking a statistical average of the minimum interface radii, for any given time and temperature of sintering it was felt that only those contact areas which were originally point contacts would be averaged.

RESULTS

Initially plots were made relating the square of the ratio $(\bar{x}/\bar{a})$ to time for a given temperature of sintering, in order to determine whether a straight line relation-

TABLE I. The volume diffusion coefficient, D_v, of copper

T°C	Time (t) hr.	$\bar{x}\times10^{-4}$ cm	$\bar{a}\times10^{-2}$ cm	$\bar{x}/\bar{a}^5$	D_v(cm³/sec.)
700	4.00	18.4	3.16	$.66\times10^{-6}$	1.7×10^{-13}
	28.00	23.6		4.0	
	50.00	28.0		5.5	
	98.00	31.4		10.0	
900	3.00	35.2	3.10	$.93\times10^{-5}$	4.6×10^{-12}
	26.25	39.4		3.44	
	50.25	48.0		8.90	
	98.25	53.6		16.00	
1000	3.00	33.0	2.84	2.1×10^{-5}	1.3×10^{-11}
	20.75	47.2		11.4	
	43.00	52.0		21.0	
	91.00	61.0		46.0	
1025	3.00	28.0	3.00	$.61\times10^{-5}$	3.5×10^{-11}
	27.00	60.0		30.0	

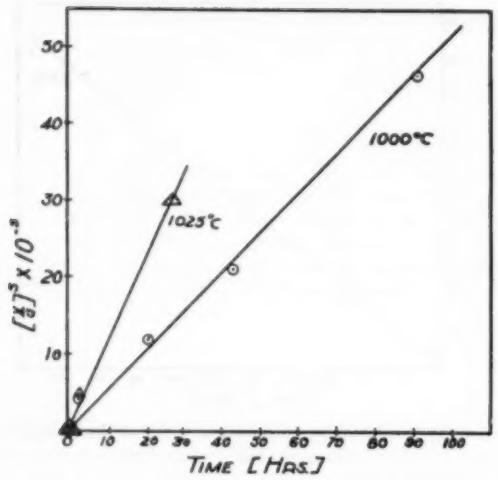


FIG. 4. Plot of the fifth power of the ratio of the average interface radius $\bar{x}$ to the average particle radius $\bar{a}$ against time. Sintering temperatures, 1000°C and 1025°C.

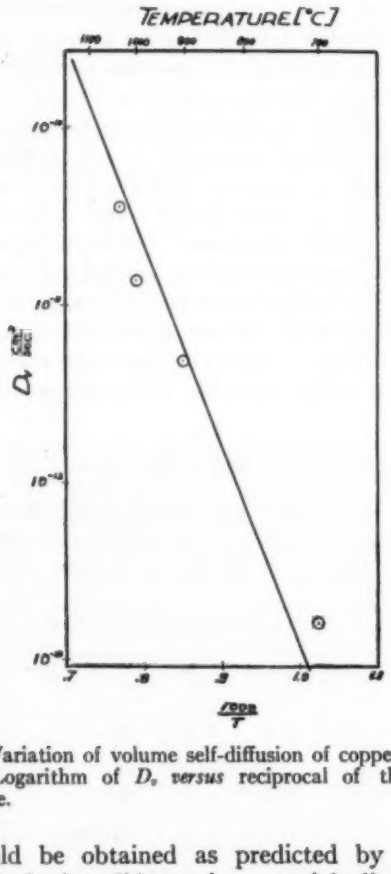


FIG. 5. Variation of volume self-diffusion of copper with temperature. Logarithm of D_v *versus* reciprocal of the absolute temperature.

ship would be obtained as predicted by Frenkel's Eq. (1). Such plots did not give a straight line relationship, the ratio of the square of $(\bar{x}/\bar{a})$ *versus* time falling off considerably with increasing times, for a given temperature of sintering.

TABLE II. The heat of activation Q_v of self-diffusion, and frequency factor D_v^0 for copper, as determined by various investigators

Authors	Q_v, cal./mole	D_v^0, cm²/sec.
F. N. Rhines and R. F. Mehl[a]	47,000	
B. V. Rollin[b]	61,000	47.0
J. Steigman, W. Shockley, and F. C. Nix[c]	57,000	0.9–11.0
C. I. Raynor, L. Thomassen, and L. J. Rouse[d]	46,800	0.3
M. S. Maier and H. R. Nelson[e]	54,000–61,000	0.1–0.6
G. Kuczynski[f]	56,000	70.0
This investigation	55,000	0.12

[a] F. N. Rhines and R. F. Mehl, Trans. A.I.M.E. **128**, 285 (1938).
[b] B. V. Rollin, Phys. Rev. **55**, 231 (1931).
[c] Steigman, Shockley, and Nix, Phys. Rev. **56**, 13 (1939).
[d] Raynor, Thomassen, and Rowe, Trans. Am. Soc. Metals, **30**, 313 (1942).
[e] M. S. Maier and H. R. Nelson, Trans. A.I.M.E. **147**, 39 (1942).
[f] See reference 3.

Next plots were made relating the fifth power of $(\bar{x}/\bar{a})$ to time for a given temperature of sintering. Such a plot, according to Kuczynski's equation for volume diffusion (2) should result in a straight line. These plots shown in Figs. 2 through 4 do show straight line relationships thus lending substantiation to Kuczynski's theory of the mechanism of sintering of crystalline bodies.

Calculation of D_v, the coefficient of volume diffusion, was made by use of Eq. (2) and the data in Table I. The use of Eq. (2) involves assuming a value of surface tension for solid copper. In this case, the value of the surface tension of liquid copper was used on the assumption that it should be very close to that for solid copper. Any error introduced by this assumption would not change the order of magnitude of the coefficient of volume diffusion calculated by means of Eq. (2). The value used for the surface tension of copper was 1200 ergs/cm². On the foregoing basis the values of D_v so obtained are given in Table I, for the various temperatures investigated in this work.

The heat of activation of volume diffusion Q_v, for copper and the frequency factor D_v^0 were calculated by plotting D_v *versus* $1/T$, as shown in Fig. 5. The value of Q_v obtained in this manner is 55,000 calories/mole while the value of D_v^0 is .12 (cm²/sec.). These values are in close agreement with those obtained by Kuczynski, and by other workers using radioactive tracer methods, as shown in Table II.

CONCLUSIONS

1. A simple, quick, non-metallographic method has been devised for studies of the rate and mechanism of sintering of crystalline bodies.

2. Strong evidence has been obtained which lends support to Kuczynski's theory that the mechanism of sintering of large crystalline particles is mainly one of volume diffusion.

Harmonic and Subharmonic Response for the Duffing Equation*

$\ddot{x}+\alpha x+\beta x^3=F\cos\omega t\ (\alpha>0)$

MORRIS E. LEVENSON
Department of Mathematics, The Cooper Union, New York, New York

(Received April 11, 1949)

The Duffing equation $\ddot{x}+\alpha x+\beta x^3=F\cos\omega t$ $(\alpha>0)$ is known to possess periodic solutions with frequency ω/n for all integral n, provided that β is sufficiently small. For $n=1$ and $n\neq 1$, these solutions are designated in the literature as harmonic and subharmonic solutions respectively. In this paper, a classification of the various types of periodic solutions is given, in which it is shown that there are two types of harmonic solutions and two types of subharmonic solutions for sufficiently small β. The perturbation method is employed to find the approximate response curves for each of the four types of periodic solutions. A comparison is made between the response curves obtained in the non-linear case $(\beta\neq 0)$ and linear case $(\beta=0)$ and some properties of the solutions in the non-linear case are discussed. A comparison is then made between the perturbation method and the Rauscher method, which is an iteration method that assumes F small instead of β. This comparison is not made in all generality, but only for one particular type of subharmonic solution. The main result obtained is that for F small, the two methods yield similar results for larger values of β than might have been anticipated.

1. INTRODUCTION

IT is known[1] that the Duffing [8]** equation $\ddot{x}+\alpha x+\beta x^3=F\cos\omega t$ possesses periodic solutions with frequency equal to $1/n$th of the frequency ω of the impressed force for all integral n, and that these solutions and corresponding frequencies are analytic functions of β for arbitrary F provided that β is sufficiently small. The cases $n=1$ and $n\neq 1$ are designated in the literature [10, 11, 17] as the harmonic and subharmonic cases respectively. It is the purpose of this paper to show firstly, how a classification of the various types of harmonic and subharmonic solutions for sufficiently small β may be obtained and some of the properties of these solutions anticipated, by a simple analysis of the linear equation $\beta=0$, and secondly, to compare our results with those obtained by the Rauscher [29] method, in which approximate solutions are obtained under the assumption that F instead of β is small. This comparison is not carried out in all generality, but only for the subharmonics of order 3 $(n=3)$.

Before we consider these questions, we wish to make the following remarks[2] that are known concerning the Duffing equation without an external force—the free non-linear equation $(F=0)$. For $F=0$, a solution can be obtained by direct integration in terms of elliptic integrals. A qualitative study may also be made by investigating the nature of the singular points and the topological character of the trajectories in the phase plane, that is, the $(x, \dot{x})$ plane, identifying the singular points with points of equilibrium and the closed trajectories with periodic solutions. It is found that when β is positive, all solutions are periodic and when β is negative, periodic solutions will exist only when the maximum displacement $|a|$ satisfies the inequality $|a|\leq(\alpha/|\beta|)^{\frac{1}{2}}$, the equality sign giving the position of unstable equilibrium. It is also found that the frequency of the solution depends upon the amplitude, a state of affairs which is quite different for the linear case, $\beta=0$, in which the frequency remains fixed for all values of the amplitude.

2. THE NON-LINEAR EQUATION

Consider the Duffing equation in the non-dimensional form

$$\nu^2\xi''+\xi+\delta\xi^3=\cos\theta \tag{1}$$

under the initial conditions

$$\xi(0)=M;\quad \xi'(0)=0 \tag{2}$$

where the differentiation refers to the variable $\theta=\omega t$, and where

$$\nu^2=\omega^2/\alpha,\quad \delta=\beta F^2/\alpha^3,\quad M=A\alpha/F,\quad \xi=\alpha x/F. \tag{3}$$

For $\delta=0$, that is, for the linear case, the solutions are given by

$$\xi=\left(M+\frac{1}{\nu^2-1}\right)\cos\frac{\theta}{\nu}-\frac{1}{\nu^2-1}\cos\theta \tag{4}$$

The various types of periodic solutions of the linear equation may be classified as follows.

(a) For $\nu=(1-1/M)^{\frac{1}{2}}(M\neq 1)$, the solutions are periodic with period 2π in θ and are called "ordinary harmonic" solutions. The frequency[3] of the solutions is the same as the frequency of the exciting force.

(b) For $\nu=n$ (n, a positive integer $\neq 1$), the solutions are periodic with period $2n\pi$ in θ, and are called "ordinary subharmonic" solutions.

---

* The present paper is an abridgment of the author's doctoral dissertation, New York University, 1948. See reference [20] in the bibliography.

[1] See in particular references [9, 20] in the Bibliography. See also the existence proofs of Poincaré [28] and Mandelstam and Papalexi [23].

** Numbers in brackets denote references in the Bibliography.

[2] These results may be found in the following references [8, 10, 11, 30, 34]. For discussions on singular points and trajectories in the phase plane associated with free systems, see the following [4, 13, 26].

[3] All frequencies discussed are circular frequencies.

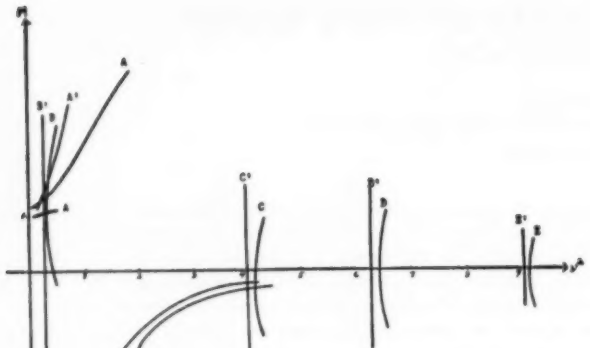


FIG. 1. Approximate response curves for the various types of periodic solutions, with $\delta=0.1$. Superimposed for comparison are the corresponding exact curves for $\delta=0$.

A	Ordinary harmonic $\delta=0.1$
A'	Ordinary harmonic $\delta=0$
B	Ultraharmonic $r/s=\frac{1}{2}$; $\delta=0.1$
B'	Ultraharmonic $r/s=\frac{1}{2}$; $\delta=0$
C	Ordinary subharmonic $r/s=2$; $\delta=0.1$
C'	Ordinary subharmonic $r/s=2$; $\delta=0$
D	Ultra-subharmonic $r/s=5/2$; $\delta=0.1$
D'	Ultra-subharmonic $r/s=5/2$; $\delta=0$
E	Ordinary subharmonic $r/s=3$; $\delta=0.1$
E'	Ordinary subharmonic $r/s=3$; $\delta=0$

(c) For $\nu=1/m$ (m, a positive integer $\neq 1$), the solutions are periodic with period 2π in θ, and are called "ultraharmonic" solutions, to distinguish them from the ordinary harmonics which have the same property, that is, they have the same frequency as the exciting force.

(d) For $\nu=p/q$ (p and q, integers relatively prime and $p, q\neq 1$), the solutions are periodic with period $2p\pi$ in θ, and are called "ultra-subharmonic" solutions, to distinguish them from the ordinary subharmonics which have the same property, that is, their frequency is an integral submultiple of the frequency of the exciting force.

It is plausible to expect that in the non-linear case different types of periodic solutions will be found whenever the non-dimensional frequency ν lies in the neighborhood of each of the values $(1-1/M)^{\frac{1}{2}}$, n, $1/m$, and p/q. This is indeed the case. The existence of the various types of periodic solutions to be discussed may be proved by the methods employed in the references cited above. These proofs will be omitted here. We now classify the periodic solutions of the non-linear Duffing equation as follows.

The "ordinary harmonic" solutions of Eq. (1) under the initial conditions (2) are those periodic solutions of frequency 1 with the property that ν reduces to $(1-1/M)^{\frac{1}{2}}$ when $\delta\to 0$.

The "ordinary subharmonic" solutions are those periodic solutions of frequency $1/n$ with the property that ν reduces to n when $\delta\to 0$ (n, a positive integer $\neq 1$).

The "ultraharmonic" solutions are those periodic solutions of frequency 1 with the property that ν reduces to $1/m$ when $\delta\to 0$ (m, a positive integer $\neq 1$).

The "ultra-subharmonic" solutions are those periodic solutions of frequency $1/p$ with the property that ν reduces to p/q when $\delta\to 0$ (p and q, integers relatively prime and $\neq 1$).

Perhaps the simplest method for obtaining approximate periodic solutions of the Duffing equation is the perturbation method.[4] Briefly, the procedure is the

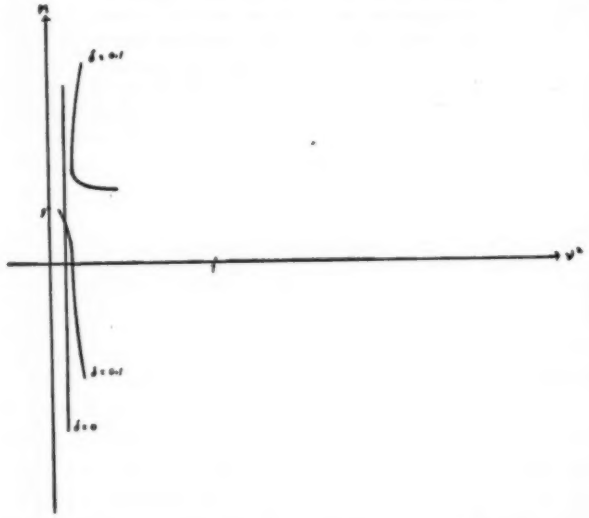


FIG. 2. Approximate response curve for the ultraharmonic solution ($r/s=\frac{1}{2}$) with $\delta=0.1$. Superimposed for comparison is the corresponding exact curve for $\delta=0$.

following. Assume the developments in powers of δ

$$\xi=\xi_0(\theta)+\xi_1(\theta)\delta+\xi_2(\theta)\delta^2+\cdots, \tag{5}$$

$$\nu=\nu_0+\nu_1\delta+\nu_2\delta^2+\cdots \tag{6}$$

with the initial conditions

$$\xi_0(0)=M;\quad \xi_i(0)=0,\quad (i=1, 2\cdots);\qquad \xi_i'(0)=0,\quad (i=0, 1, 2\cdots) \tag{7}$$

and the periodicity conditions

$$\xi_i(\theta+\lambda)=\xi_i(\theta),\quad (i=0, 1, 2\cdots) \tag{8}$$

where λ is to have either of the values 2π or $2\sigma\pi$ (σ, an integer $\neq 1$) depending upon the type of periodic solution we are looking for. Note that in this procedure the initial value M is prescribed and the frequency ν is determined as a function of M.

The series (5) and (6) are now inserted in the differential equation (1) and like powers of δ are equated. This yields a set of linear differential equations

$$\nu_0^2\xi_0''+\xi_0=\cos\theta, \tag{9}$$

$$\nu_0^2\xi_1''+\xi_1=-\xi_0^3-2\nu_0\nu_1\xi_0'', \tag{10}$$

. . . . . .

. . . . . .

[4] For justification of this procedure see [9, 20].

Equation (9) is solved for ξ_0 and we obtain

$$\xi_0=\left(M+\frac{1}{\nu_0^2-1}\right)\cos\frac{\theta}{\nu_0}-\frac{1}{\nu_0^2-1}\cos\theta. \qquad (11)$$

The type of periodic solution we are looking for depends upon the choice of ν_0 in this solution and this in turn fixes the λ in Eq. (8). Let us choose $\nu_0=r/s$ where r and s are integers relatively prime.

For $r/s=n$, the solution $\xi(\theta)$ to be determined is an ordinary subharmonic solution of period $2n\pi$ ($\lambda=2n\pi$).

For $r/s=1/m$ the solution $\xi(\theta)$ to be determined is an ultraharmonic solution of period 2π ($\lambda=2\pi$).

For $r/s=p/q$ where p and q are different from unity, the solution $\xi(\theta)$ to be determined is an ultrasubharmonic solution of period $2p\pi$ ($\lambda=2p\pi$).

Inserting ξ_0 and ν_0 in the right side of Eq. (10) and making use of the identities

$$\cos^3\alpha=\tfrac{3}{4}\cos\alpha+\tfrac{1}{4}\cos 3\alpha,$$

$$\cos^2\alpha=\frac{1+\cos 2\alpha}{2},$$

$$\cos\alpha\cos\beta=\frac{\cos(\alpha+\beta)+\cos(\alpha-\beta)}{2},$$

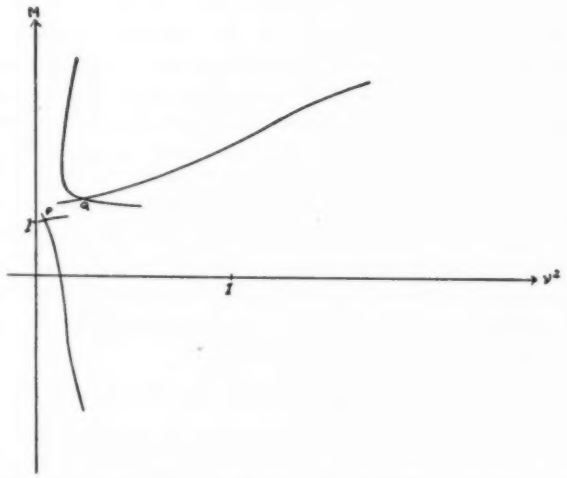


FIG. 3. Superposition of the approximate ordinary harmonic and ultraharmonic ($r/s=\frac{1}{3}$) response curves, with $\delta=0.1$; showing intersections P and Q. The branch of the ordinary harmonic curve below the ν^2 axis is omitted from the figure.

we set the coefficient of the term $\cos(s/r)\theta$ in the right side of Eq. (11) equal to zero to satisfy the periodicity condition for ξ_1. This will determine ν_1. We thus obtain the following approximate response relations:

For $r/s\neq 3, \frac{1}{3}$

$$\nu^2=(r/s)^2\left\{1+\tfrac{3}{4}\delta\left[M^2+\frac{2M}{(r/s)^2-1}+\frac{3}{\{(r/s)^2-1\}^2}\right]\right\}+\cdots. \qquad (12)$$

For $r/s=3$

$$\nu^2=9\left\{1+\frac{3}{4}\delta\left[M^2+\frac{M}{8}+\frac{1}{32}\right]\right\}+\cdots. \qquad (13)$$

For $r/s=\frac{1}{3}$

$$\nu^2=\frac{1}{9}\left\{1+\frac{3}{4}\delta\left[M^2-\frac{9M}{4}+\frac{243}{8}\left(\frac{M-1}{8M-9}\right)\right]\right\}+\cdots. \qquad (14)$$

The exceptional character of the cases $r/s=3, \frac{1}{3}$ are essentially due to the presence of the cubic term in the restoring force.

In order to obtain the approximate response relation for the ordinary harmonic case where ν_0 in Eq. (11) has the value $(1-1/M)^{\frac{1}{2}}$, the above procedure must be slightly modified. After inserting the value $\xi_0=M\cos\theta$ and $\nu_0=(1-1/M)^{\frac{1}{2}}$ in the right side of Eq. (10), we solve for ξ_1 and then set the coefficient of the term $\cos[\theta/(1-1/M)^{\frac{1}{2}}]$ equal to zero in this solution. This will determine ν_1. The following approximate response relation is obtained

$$\nu^2=1-\frac{1}{M}+\delta M^2\left(\frac{6M-7}{8M-9}\right)+\cdots. \qquad (15)$$

In order to interpret the above approximate response relations correctly, it is essential that we give a precise statement of the existence theorems.

The ordinary harmonic, ordinary subharmonic, ultraharmonic, and ultra-subharmonic solutions of Eq. (1) under the initial conditions (2) exist for sufficiently small δ and these solutions and their corresponding frequencies ν are analytic in the region $|M|<M_1$, $|\delta|<\delta_1$ where M_1 is arbitrary but finite and where δ_1 is sufficiently small. There is one exception; for the ordinary harmonic case and the ultraharmonic case $r/s=\frac{1}{3}$ the existence proof breaks down for the initial value $M=9/8$.

Figures 1 and 2 show plots in the (ν^2, M) plane of the

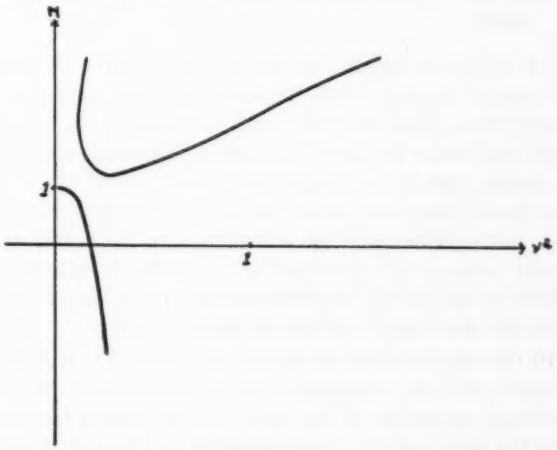


FIG. 4. Conjectured exact response curves for the ordinary harmonic and ultraharmonic solution ($r/s=\frac{1}{3}$), with $\delta=0.1$. The branch of the ordinary harmonic curve below the ν^2 axis is omitted from the figure.

approximate response relations (12), (13), (14) and (15) for several values of r/s and for $\delta=0.1$. Superimposed for comparison are the corresponding exact curves for the linear case, $\delta=0$. In the non-linear case, the curves may be thought of as arising from the curves in the linear case by bending the latter to the right. For $\delta<0$ the curves bend to the left—we omit the figure for this case. A marked deviation occurs between the ordinary harmonic curves for $\delta=0.1$ and $\delta=0$ and between the ultraharmonic curves $r/s=\frac{1}{3}$ for $\delta=0.1$ and $\delta=0$ in the neighborhood of $M=9/8$. The curves for $\delta=0$ in each case split into two branches there to produce the corresponding curves for $\delta=0.1$. It should be emphasized that $M=9/8$ is just the initial value for which the existence theorem breaks down in both cases, so that it is questionable as to what the true nature of the curves are for $\delta=0.1$ in the neighborhood of $M=9/8$. We shall return to this point later.

Since every point of the exact response curves which correspond to the approximate curves of Fig. 1 represents a periodic solution of the Duffing equation (1) under the initial conditions (2), one sees how numerous the periodic solutions of this equation are for fixed but sufficiently small δ. More precisely, consider the region $|M|<M_1$, $0<\nu^2<\nu_1^2$ where ν_1 and M_1 are finite but arbitrary, and cover this region with a net of squares with side ϵ where ϵ is small. Consider the N_1 squares in the interior of the given region and let the point (ν_k^{*2}, M_k^*) denote the center of the kth square. The following theorem is easily established. For the proof of this theorem, see [20].

Theorem: There exists a ν^2 and an M satisfying the inequalities

$$|\nu^2-\nu_k^{*2}|<\epsilon/2, \quad |M-M_k^*|<\epsilon/2, \qquad k=1, 2, \cdots N_1(\epsilon)$$

where ϵ is small, for which the differential equation (1) under the initial conditions (2) possesses periodic solutions, provided that δ is sufficiently small.

Since ϵ may be arbitrarily small in the above theorem, one might suspect that the set of periodic solutions is everywhere dense in every finite portion of the (ν^2, M) right half-plane for fixed but sufficiently small δ. This is of course true in the linear case, $\delta=0$; however, in the non-linear case, we cannot make this assertion because it is not known whether a fixed δ can be found that would insure the existence of periodic solutions for infinitely many r/s simultaneously. It is conjectured, however, that such a δ can be found.

In the approximate relations (12) and (13) it will be noticed that the coefficients of δ are positive definite quadratic forms in M for all r/s. Combining this fact with the fact that the ν's are analytic in δ and M in some region $|\delta|<\delta_1$, $|M|<M_1$, we can choose δ small enough to make the exact value of ν for $\delta>0$ in each of the above cases greater than its value for $\delta=0$. When $\delta<0$, the exact value of ν may be made less than its value for $\delta=0$. This is significant, since it states that no ordinary subharmonic solution exists whose frequency is exactly equal to the free linear frequency. Some writers define subharmonic response as an oscillation with a frequency equal to the frequency of the free linear oscillation excited by a force whose frequency is an integral multiple of the free linear frequency. At least for the Duffing equation, this cannot occur.

We conclude this section with a few remarks concerning the question of "bifurcation" defined below. Let us assume for the moment that the approximate response curves in Fig. 1 are the exact response curves. The ordinary harmonic curve for $\delta=0.1$ intersects each of the other types of response curves for $\delta=0.1$. At these points of intersection we have a fixed M, ν and δ, and since $\xi'(0)=0$ is prescribed, these points correspond to unique periodic solutions of the Duffing equation. Such points of intersection are called "bifurcation" points, or, bifurcation points correspond to such values of the parameters of the differential equation for which two different types of periodic solutions become identical. In the linear case, the ordinary harmonic solution bifurcates uniquely with each of the types of periodic solutions. It is easy to show (see [20]) that in the non-linear case the same is true provided that δ is small enough. There is one exception in the non-linear case; we have not been able to prove bifurcation of the ordinary harmonic solution with the ultraharmonic solution $r/s=\frac{1}{3}$ because of the lack of analyticity of both solutions at $M=9/8$. In Fig. 3 we have superimposed the approximate response curves for the ordinary harmonic and ultraharmonic $(r/s=\frac{1}{3})$ cases for $\delta=0.1$. Two intersections P and Q appear. It is conjectured that the exact response curves for these two solutions merge with each other in the neighborhood of these points and the branches of the approximate curves asymptotic to the line $M=9/8$ do not appear. Figure 4 shows the conjectured exact response curves for these two solutions.

3. THE RAUSCHER METHOD

Consider the Duffing equation in its original form

$$\ddot{x}+\alpha x+\beta x^3=F\cos\omega t, \quad (\alpha>0) \tag{16}$$

with $x(0)=A$; $\dot{x}(0)=0$. The perturbation method employed in the previous section was an approximation method which started with the solution of the linear equation $\beta=0$ as a first approximation. We might expect more rapid convergence if we took for our first approximation the solution of the free non-linear equation $F=0$ and then performed iterations in the neighborhood of $F=0$. This is what Rauscher [29] has done. He assumes the solution to be monotonic in the half-period and uses graphical methods due to the presence of complicated hyper-elliptic integrals. Briefly his method is the following. We first set $\theta=\omega t$ in the differential equation (16) to work with a fixed period, and obtain

$$\omega^2 x''+\alpha x+\beta x^3=F\cos\theta \tag{17}$$

with $x(0)=A$; $x'(0)=0$, where the differentiation is with respect to θ. We look for periodic solutions of period $2n\pi$. For a first approximation, we take the solution of

$$\omega_0^2x''+\alpha x+\beta x^3=0 \tag{18}$$

which is obtained by explicit integration as follows. Setting $v=dx/d\theta$ and negative in the interval $0<\theta<n\pi$, Eq. (18) reduces to

$$\omega_0^2(v^2/2)+G(x)-G(A)=0 \tag{19}$$

where $G(x)=\alpha(x^2/2)+\beta(x^4/4)$. Equation (19) reduces to

$$-d\theta_0=\frac{\omega_0}{\sqrt{2}}\frac{dx}{(G(A)-G(x))^{\frac{1}{2}}}, \tag{20}$$

or

$$\theta_0(x)=\frac{\omega_0}{\sqrt{2}}\int_x^A\frac{dx}{(G(A)-G(x))^{\frac{1}{2}}}, \tag{21}$$

and ω_0 is obtained from Eq. (20) by integration over the quarter-period

$$\frac{\omega_0}{\sqrt{2}}=\frac{\frac{n\pi}{2}}{\int_0^A\frac{dx}{(G(A)-G(x))^{\frac{1}{2}}}}. \tag{22}$$

To find the next approximation, the solution $\theta_0(x)$ is inserted in the right side of Eq. (17) resulting in the equation

$$\omega_1^2x''+\alpha x+\beta x^3-F\cos\theta_0(x)=0 \tag{23}$$

from which is obtained, by the same procedure as above

$$\theta_1(x)=\frac{\omega_1}{\sqrt{2}}\int_x^A\frac{dx}{\left(G(A)-G(x)-F\int_x^A\cos\theta_0(x)dx\right)^{\frac{1}{2}}} \tag{24}$$

and

$$\frac{\omega_1}{\sqrt{2}}=\frac{\frac{n\pi}{2}}{\int_0^A\frac{dx}{\left(G(A)-G(x)-F\int_x^A\cos\theta_0(x)dx\right)^{\frac{1}{2}}}}. \tag{25}$$

The last two relations assume that n is odd. When n is even ω_1 would be obtained by integrating over the half-period

$$\frac{\omega_1}{\sqrt{2}}=\frac{n\pi}{\int_{-B}^A\frac{dx}{\left(G(A)-G(x)-F\int_x^A\cos\theta_0(x)dx\right)^{\frac{1}{2}}}} \tag{26}$$

where $-B$ is the value of x for $\theta_1=n\pi$ and is determined from the assumption that the work done by the restoring force $\alpha x+\beta x^3-F\cos\theta_0(x)$ in the intervals $0\leq x\leq A$ and $0\geq x\geq -B$ is the same; that is

$$\int_0^A[\alpha x+\beta x^3-F\cos\theta_0(x)]dx=\int_0^{-B}[\alpha x+\beta x^3-F\cos\theta_0(x)]dx. \tag{27}$$

The iteration procedure is now evident. The nth approximation is obtained from

$$\omega_{n-1}^2x''+\alpha x+\beta x^3-F\cos\theta_{n-2}(x)=0. \tag{28}$$

We can obtain some idea of the behavior of the frequency of the solution in this iteration procedure by avoiding the graphical method and applying the following short-cut method suggested by Rauscher. The use of the short-cut method rather than the true method will be justified later. We consider the case $n=3$. As a first approximation, we take $x=A\cos(\theta/3)$, the solution of the free linear equation with $\omega_0=3\sqrt{\alpha}$. We then write this approximation in the form $\cos(\theta_0/3)=x/A$ and obtain the following relation for $\cos\theta_0(x)$.

$$\cos\theta_0(x)=4(x^3/A^3)-3(x/A) \tag{29}$$

where we have made use of the identity $\cos\alpha=4\cos^3(\alpha/3)-3\cos(\alpha/3)$. This value of $\cos\theta_0(x)$ is then substituted in the right side of Eq. (17) and the next approximation is obtained from

$$\omega_1^2x''+(\alpha+3(F/A))x+(\beta-4(F/A^3))x^3=0. \tag{30}$$

The frequency ω_1 is now obtained by direct integration and it is given by

$$\omega_1/\sqrt{2}=\frac{\frac{3\pi}{2}}{\int_0^A\frac{dx}{(R(A)-R(x))^{\frac{1}{2}}}} \tag{31}$$

where

$$R(x)=\left(\alpha+3\frac{F}{A}\right)\frac{x^2}{2}+\left(\beta-4\frac{F}{A^3}\right)\frac{x^4}{4}.$$

The integral appearing in Eq. (31) can be transformed into a complete elliptic integral of the first kind and we obtain

$$\frac{\omega_1}{\sqrt{\alpha}}=\nu=\frac{3\pi(1+\mu-G)^{\frac{1}{2}}/2}{K\bmod k}\quad\text{for}\quad\mu\geq 4G \tag{32}$$

where $\mu=\beta A^2/\alpha$, $G=F/A\alpha$ and K is the complete elliptic integral of the first kind with modulus k and

where the modulus is given by

$$k^2=\frac{(\mu/2)-2G}{1+\mu-G}, \tag{33}$$

$$\frac{\omega_1}{\sqrt{\alpha}}=\nu=\frac{3\pi(1+\mu/2+G)^{\frac{1}{2}}/2}{K' \bmod k'} \quad \text{for} \quad \mu\leq 4G \tag{34}$$

where K' is the complete elliptic integral of the first kind with modulus k' and where

$$k'^2=\frac{2G-\mu/2}{1+\mu/2+G}. \tag{35}$$

It is interesting to note that the parameters $\beta A^2/\alpha=\mu$, $F/A\alpha=G$ and $\omega/\sqrt{\alpha}=\nu$ are the natural non-dimensional parameters in this iteration scheme. That this should be so can be seen, if in our differential equation (17) we set $\xi=x/A$ and divide through by αA. This yields the differential equation in the non-dimensional form

$$\nu^2\xi''+\xi+\mu\xi^3=G\cos\theta \tag{36}$$

with the initial conditions $\xi(0)=1$; $\xi'(0)=0$.

Figure 5 shows a plot in the $\mu-G$ plane, of the curves $\omega^2/\alpha=\nu^2=$ constant, determined from the relations (32), (33), (34), and (35). For purposes of comparison we have included in the same figure the curves $\omega^2/\alpha=\nu^2=$ constant obtained from the perturbation method of the previous section, that is, we use the Eq. (13) after transforming it in terms of the new non-dimensional parameters. Equation (13) then reduces to

$$\omega^2/\alpha=\nu^2=9\{1+\mu[\tfrac{3}{4}+\tfrac{3}{32}G+(3/128)G^2]\}+\cdots.$$

The smooth curves in Fig. 5 are those obtained by the Rauscher short-cut method and the short-dashed curves are those obtained by the perturbation method. The most significant result, on comparing the two sets of curves, is that there is excellent agreement for larger values of μ than might have been anticipated. However, for values of μ in the neighborhood of -1 the agreement is not as good. This is due to the fact that the perturbation curves do not describe the situation accurately for such values of μ since we know that when $\mu=-1$ and $G=0$, we should have a point of unstable equilibrium or $\nu^2=0$ (see Introduction).

How can we account for the more apparent discrepancy, namely, the fact that the Rauscher curves are limited to the angular region bounded by the lines $\nu^2=0$? And how can we account for the fact that when $\mu=0$ the Rauscher curves show that ν^2 may have any value between 0 and 9, while the perturbation curves yield the constant value $\nu^2=9$? We offer two reasonable explanations for these apparent contradictions. First, the Rauscher method assumes the solution to be monotonic in the half-period $0\leq\theta\leq 3\pi$, while there is no such restriction in the perturbation method. In fact, it can be shown that for $\mu=0$, that is, the linear equation, the solution is not monotonic in the above interval when $G>1$ and when $G<-2$. Second, the approximation obtained by the Rauscher method may be a bad one for very small μ even when G is small. In other words, The Rauscher approximation may only be good for such small values of G which vanish with μ.

The short-cut method which we used instead of the true Rauscher method did not begin in the neighborhood of the free non-linear equation so that the reliability of our Rauscher curves might be open to question. However, the true Rauscher method was also carried out for several values of the parameters μ and G, and satisfactory agreement was found between the two methods. In particular, computations were carried out for $\mu=-0.9$, $\mu=0.2$, and $\mu=2$ and for values of G in the range $-1\leq G\leq 2$. It was anticipated that there would be satisfactory agreement for small G, but it was surprising to find that for G not small, the agreement was very good. An analysis of the true Rauscher method also revealed that the two methods yielded exactly the same

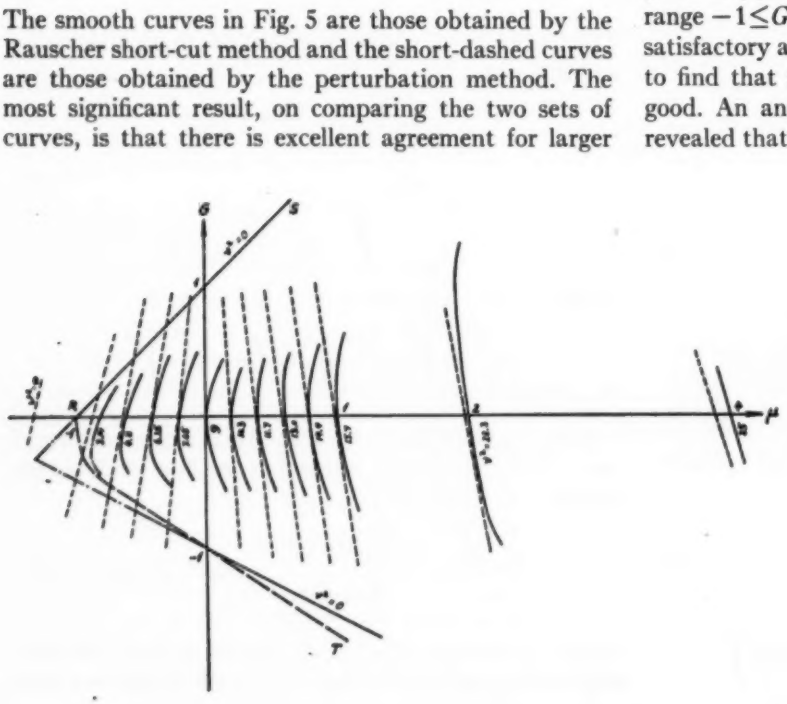


FIG. 5. Comparison between the approximate response curves obtained by the Rauscher short-cut method and perturbation method, for the ordinary subharmonic solution ($r/s=3$).

- — — — — perturbation method
- ———— short-cut Rauscher method
- —— —— correction of boundary curve $\nu^2=0$ obtained by the true Rauscher method
- —·—·— bad approximation obtained by the short-cut Rauscher method

results along the line $G=0$ for all $\mu \geq -1$ and along the line $\mu=0$ for all G in the interval $-1 \leq G \leq 1$. The analysis also revealed that the true Rauscher boundary curves $\nu^2=0$ again define an angular region, this time with the angle at the unstable equilibrium point $\mu=-1$, $G=0$. The agreement between the two methods is therefore bad in the neighborhood of $\mu=-1$ and $G<0$. In Fig. 5, the true boundary curve $\nu^2=0$ consists of the line RS which incidentally coincides with part of the short-cut boundary curve, and the curve RT, indicated by the long dashes. Those portions of the short-cut Rauscher curves which are bad approximations are indicated in the figure by dot-dash segments.

In Fig. 6, we show a plot in the $\nu-G$ plane of the true Rauscher response curves for $\mu=0, 0.2, 2$. Superimposed for comparison are the corresponding perturbation response curves.

The author wishes to express his thanks to Professors K. O. Friedrichs and J. J. Stoker of New York University for their many valuable suggestions and helpful criticisms.

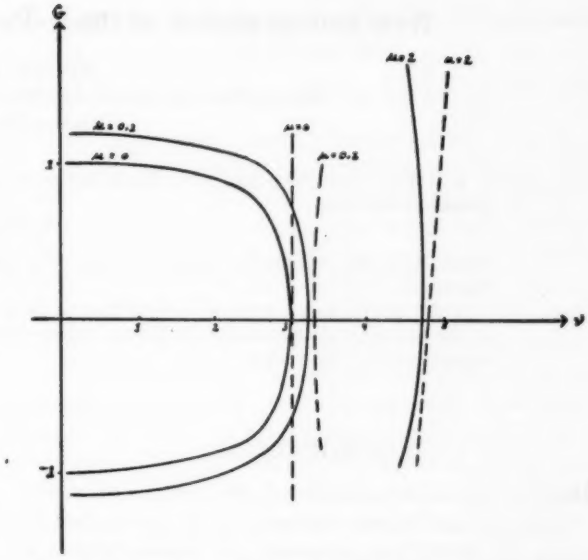


FIG. 6. Comparison between the approximate response curves obtained by the true Rauscher method and perturbation method, for the ordinary subharmonic solution ($r/s=3$).
– – – – perturbation method
——— true Rauscher method

BIBLIOGRAPHY

(1) E. V. Appleton, "On the anomolous behavior of a galvanometer," Phil. Mag., Series 6, **47**, 609 (March 1924).
(2) E. V. Appleton and W. M. H. Greaves, "On the solution of the triode oscillator," Phil. Mag., Series 6, **45**, 401 (March 1923).
(3) J. G. Baker, "Forced vibrations with non-linear spring constants," Trans. Am. Soc. Mech. Eng. **54** (1932).
(4) L. Bieberbach, *Differentialgleichungen* (Verlag, Julius Springer, Berlin, 1930), third edition, Part I, Chapter IV.
(5) J. P. Den Hartog, *Mechanical Vibrations* (McGraw-Hill Book Company, Inc., New York, 1940), second edition, Chapters VII and VIII.
(6) J. P. Den Hartog, "The amplitudes of non-harmonic vibrations," J. Frank. Inst. **216**, No. 4 (October 1933).
(7) J. P. Den Hartog, "Forced vibrations with non-linear spring constants," Trans. Am. Soc. Mech. Eng. **54**, 157 (1932).
(8) G. Duffing, *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz* (Friedrich Vieweg and Sohn, Braunschweig, 1918).
(9) K. O. Friedrichs, "Justification of the perturbation method," lectures delivered at Brown University, December 5, 1942; January 23, 1943.
(10) K. O. Friedrichs and J. J. Stoker, Lectures on non-linear vibrations delivered at New York University, 1939, 1941, and 1942.
(11) K. O. Friedrichs and J. J. Stoker, "Forced vibrations of systems with non-linear restoring force," Quart. App. Math. **1**, No. 2 (July, 1943).
(12) A. Hammerstein, "Eine nichtlineare Randwertaufgabe," Jahresber. d. D. Math. Ver. **39**, 59 (1930).
(13) W. Hurewicz, Lectures on differential equations delivered at Brown University 1943, Chapter IV.
(14) R. Iglisch, "Zur theorie der Schwingungen," Monats. f. Math. u. Phys. (1 Mitt.) **37**, 325 (1930).
(15) R. Iglisch, "Zur theorie der Schwingungen," Monats. f. Math. u. Phys. (2 Mitt.) **39**, 173 (1932).
(16) L. S. Jacobsen and H. J. Jesperson, "Steady forced vibrations of single mass systems with symmetrical as well as unsymmetrical non-linear restoring elements," J. Frank. Inst. **220**, 467 (1935).
(17) F. John, "Harmonic vibrations with non-linear characteristics," Inst. for Math. and Mech. N. Y. U. (1946).
(18) T. Von Karman, "The engineer grapples with non-linear problems," Bull. Am. Math. Soc. **46**, No. 8 (August 1940).
(19) N. Kryloff and N. Bogoliuboff, *Introduction to Non-linear Mechanics* (Princeton University Press, Princeton, 1940), free translation by S. Lefschetz.
(20) M. E. Levenson, "Harmonic and subharmonic response for the Duffing equation," doctoral dissertation, New York University, 1948.
(21) A. Lindsted, *Differentialgleichungen der Störungstheorie* (Mémoires de l'Académie Impériale des Sciences de St. Petersbourg), Series 7, **31**, No. 4 (1883).
(22) C. A. Ludeke, "Resonance," J. App. Phys. **13**, 418 (1942).
(23) L. Mandelstam and N. Papalexi, "Über resonanzerscheinungen bei frequenzteilung," Zeits. f. Physik **73** (1932).
(24) L. Mandelstam and N. Papalexi, "Exposé des recherches récentes sur les oscillations non-lineaires," report presented to the radiophysics section of the International Congress of the URSI in London, September 12–18, 1934, Tech. Phys. USSR **2** (1935).
(25) J. D. McCrumm, "An experimental investigation of subharmonic currents," Trans. A.I.E.E. **60**, 533 (1941).
(26) N. Minorsky, "Introduction to non-linear mechanics," David Taylor Model Basin Report 534 (December 1944).
(27) P. O. Pederson, "Subharmonics in forced oscillations in dissipative systems," J. Acous. Soc. Am. **6**, 227–238 (1935), **7**, 64–70 (1935).
(28) H. Poincaré, *Les méthodes nouvelles de la mécanique céleste* (Paris, 1892–1893), Vol. I, Chapter III, Vol. II, Chapter IX.
(29) M. Rauscher, "Steady oscillations of systems with non-linear and unsymmetrical elasticity," J. App. Mech. **5**, No. 4 (December 1938).
(30) Lord Rayleigh, *Theory of Sound*, (MacMillan Company, Ltd. London, 1937), second edition, Vol. I, par. 67, 68b.
(31) Lord Rayleigh, "On maintenance of vibrations," Phil. Mag. Series 5, **15**, No. 94 (April 1883).
(32) J. Shohat, "On Van der Pol's and related non-linear differential equations," J. App. Phys. **15**, 568 (1944).
(33) I. K. Silverman, "On forced pseudo-harmonic vibrations," J. Frank. Inst. **217**, No. 6 (June 1934).
(34) S. Timoshenko, *Vibration Problems in Engineering* (D. Van Nostrand Company, Inc., New York, 1928), Chapter II.
(35) B. Van der Pol, "Frequency demultiplication," Nature (September 10, 1927).
(36) B. Van der Pol, "Forced oscillations in a circuit with non-linear resistance," Phil. Mag. **65**, No. 13 (January 1927).

New Interpretation of the n-Power Law in Plastic Deformation

HERBERT I. FUSFELD
Pitman-Dunn Laboratory, Frankford Arsenal, Philadelphia, Pennsylvania
(Received April 29, 1949)

It is shown that the n-power law describing a true stress-strain curve, normally written as $S=k\delta^n$, can be placed in the form

$$S=S_0(e\delta/n)^n$$

where S_0 is the engineering tensile strength. The application of this new expression to tensile curves is discussed.

A plot of S/S_0 as a function of n, for a fixed value of the logarithmic strain δ, provides a general description of the behavior of all materials in plastic deformation. Experimental data from the literature are in good agreement with the curves.

I. INTRODUCTION

PHYSICAL interpretation of the stress-strain relations for any given material is greatly aided by plotting so-called "true stress-strain" curves. For ordinate, the "true stress," the applied force is divided by the actual instantaneous area normal to the direction of the force, rather than the initial area used for computing stress in the conventional engineering stress-strain curve. As abscissa, a most convenient quantity is the logarithmic strain δ, defined as

$$\delta=\int_{L_0}^{L} dL/L=\log L/L_0$$

where L_0=original gauge length under observation, L=gauge length at time of measurement.

When the experimental data for the simple tensile tests are plotted in terms of the above quantities, it is found that for many, if not most, metals the relationship between the "true stress" S and the logarithmic strain δ can be expressed in the form

$$S=k\delta^n.$$

This is not valid in the elastic region or in the immediate region of zero plastic strain, but does hold over the major portion of the curve. The relation is true only for plastic strains, so that the total measured strain should be corrected by subtracting the elastic strain. However, the correction is negligible for appreciable strains. The quantity n is known as the strain-hardening exponent, while k is a constant which, from the definition above, represents the stress at unit strain.

II. APPLICATION OF n-POWER LAW TO INSTABILITY

There is an immediate useful relation obtainable from the n-power law in connection with determining the point at which local instability sets in when a material is stressed. In the simplest case, uniaxial tension, this relation can be shown as follows.

The instantaneous applied force F is given by

$$F=SA \tag{1}$$

where S=stress existing in the tensile specimen, A=instantaneous area normal to F so that

$$dF/d\delta=S(dA/d\delta)+A(dS/d\delta). \tag{2}$$

From the definition of δ, one has

$$L=L_0e^{\delta}. \tag{3}$$

But in order to satisfy the condition of constant volume in the plastic range

$$LA=L_0A_0,$$

or

$$A=(L_0/L)A_0=A_0e^{-\delta}, \tag{4}$$

hence

$$dA/d\delta=-A. \tag{5}$$

Further, if the relation between S and δ is given by the n-power law

$$S=k\delta^n \tag{6}$$

and

$$dS/d\delta=nk\delta^{n-1}=nS/\delta. \tag{7}$$

Substituting Eqs. (5) and (7) into Eq. (2), the expression for the rate of change of force with strain is given by

$$dF/d\delta=SA(n/\delta-1). \tag{8}$$

It is therefore evident that the force-strain curve reaches a maximum, i.e., local instability sets in, for a critical strain

$$\delta_c=n. \tag{9}$$

The significance of n thus lies in the observation that it represents the largest strain for which uniform deformation can proceed before instability or "necking" occurs. This holds only for uniaxial tension. The onset of instability in more complicated cases of combined stresses has been discussed in a previous report.[1]

III. FURTHER IMPLICATIONS OF THE n-POWER LAW

Most analyses involving the n-power law have been concerned with the relations between the strain-hardening exponent n and the critical points of instability, such as the above example. Some interesting relations can be

[1] Herbert I. Fusfeld, "Present theory of stability of plastic flow," Frankford Arsenal Ordnance Laboratory Report R-820, (August 1947).

easily derived, however, with regard to the stresses. The present paper will be confined only to the case of simple tension.

A. Relation between Constant k and Tensile Strength

In the expression $S=k\delta^n$, the quantity k is, by definition, the stress at unit strain. Since this is not a commonly used property of materials, it seems more suitable to relate k to a more familiar stress quantity, namely, the engineering tensile strength S_0. This can be done in the following manner.

The tensile strength S_0 is the ratio of the force at the point of instability to the original area A_0. The true stress at this point, S_c, is the ratio of that force to the instantaneous area of the specimen A. If the strain at instability is δ_c, these definitions can be expressed as maximum force

$$F_c=S_0A_0=S_cA_c, \tag{10}$$

but

$$A_c=A_0e^{-\delta_c}, \tag{11}$$

therefore

$$S_c=S_0e^{\delta_c}. \tag{12}$$

If the n-power law is assumed, one has the further expression for S_c, namely

$$S_c=k\delta_c^n. \tag{13}$$

However, the preceding analysis of Section II shows that in simple tension the critical strain at instability is given by $\delta_c=n$. Utilizing this, and combining Eqs. (12) and (13), there results

$$S_c=kn^n=S_0e^n \tag{14}$$

so that

$$k=S_0(e/n)^n. \tag{15}$$

B. Generalized Family of True Stress-Strain Curves

Utilizing the preceding expression for k given by Eq. (15), the n-power law can now be expressed in the form

$$S=k\delta^n=S_0(e\delta/n)^n. \tag{16}$$

To generalize this relation, it is convenient to express the true stress in units of the tensile strength. Let the quantity x be defined as the ratio

$$x=S/S_0. \tag{17}$$

Then Eq. (17) becomes

$$x=(e\delta/n)^n \tag{18}$$

where all quantities are now dimensionless.

It is interesting to discuss the ratio x as a function of n. If the strain δ is considered constant, this function has a maximum at $n=\delta$. The value of x at this maximum is, from Eq. (18), e^δ. Hence, a family of curves can be drawn for different fixed values of δ as given in Fig. 1. All the maximum points lie on the broken curve $x=e^\delta$.

C. Limiting Feature of Conventional Expression of n-Power Law

At $n=0$, the function is unity, since $\lim_{n\to 0} n^n=1$. Thus all the curves pass through the point $n=0$, $x=1$, or $S=S_0$. This is simply a mathematical statement of the physical interpretation that n represents the amount of uniform plastic strain that occurs before instability sets in. When $n=0$ no prior plastic flow takes place, the area normal to the applied force is unchanged from its initial value, and hence the true stress at the point of instability is numerically equal to the engineering tensile stress.

As illustrated by the above extreme case, the revised expression for the n-power law has the advantage of indicating the effect of changes in n. The original expression $S=k\delta^n$ is simply a description of the variation of S with δ for a specific material having a particular value of n. It was not intended to bring out the effects of varying the strain hardening exponent. Actually, of course, as n varies, the stress required to produce unit strain, i.e., k, changes in accordance with Eq. (15).

The question might arise as to the relation between the tensile strength S_0 and n. This, however, is a misleading question. More properly, one should inquire as

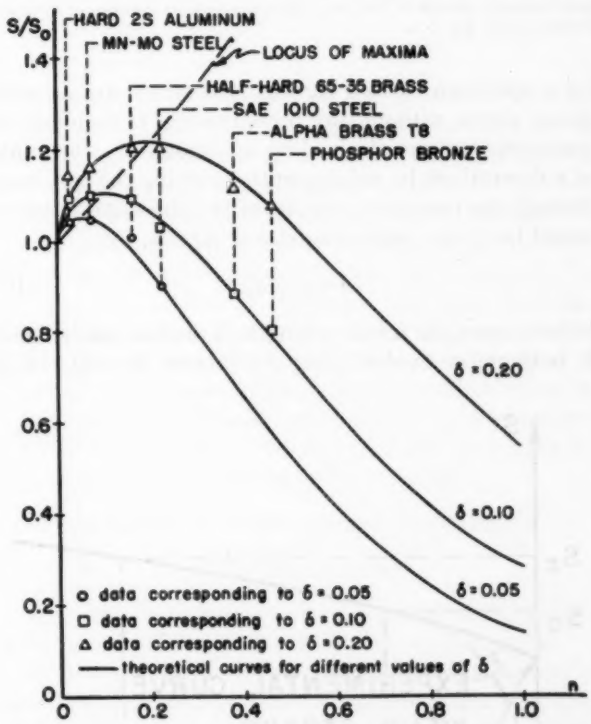


FIG. 1. Comparison of experimental data with theoretical curves of S/S_0 vs. n for strain values of 0.05, 0.10 and 0.20.

to the relation between the tensile strength of a material in a particular state and the process by which this state is changed to produce a change in n.

Consider, for example, the simple and important case of strain-hardening produced by carrying a specimen

through a small amount of uniform plastic strain, releasing the stress, then repeating the process in discrete steps until necking occurs. On an engineering stress-strain diagram, the process would be shown approximately by Fig. 2.

The points $\delta_1, \delta_2, \cdots$ represent the condition of the specimen after successive steps in the process. After each step, the amount of uniform plastic deformation that can still be obtained is decreased, thus n is decreased. Specifically, let $n=n_0$ be the value determined

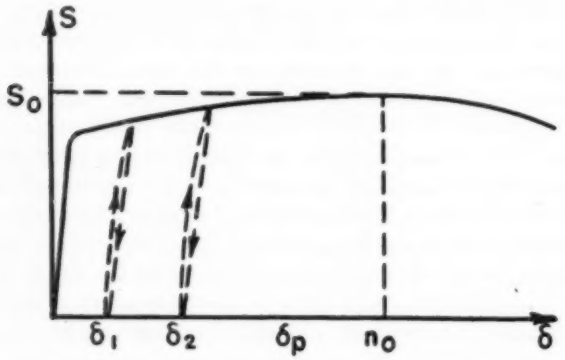


FIG. 2. Typical engineering stress-strain diagram showing the approximate course of the test when stopped at successive residual strains of $\delta_1, \delta_2, \cdots$.

for a specimen carried through the stress-strain curve shown above without any pre-straining. If an identical specimen were pre-strained by an amount δ_p, the value of n determined by subsequently carrying the specimen through the remaining portion of the stress-strain curve would be given, neglecting elastic strains, by

$$n=n_0-\delta_p. \tag{19}$$

In both cases, the tensile strength S_0 remains unchanged. It is therefore evident that the process by which n is varied is the important factor in affecting S_0, not the value of n itself.

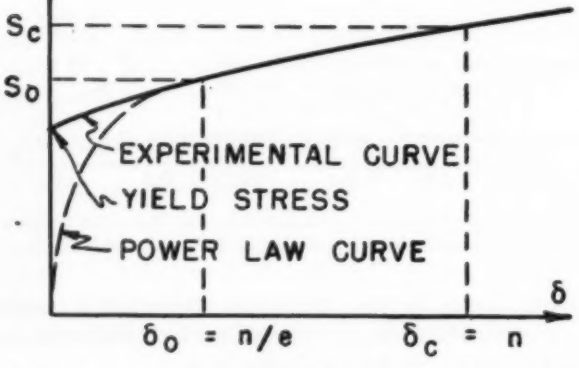


FIG. 3. True stress-strain curve showing values of strain corresponding to S_c, the true stress at instability, and S_0, the true stress numerically equal to the tensile strength.

D. Relation between Tensile Strength and True Stress-Strain Curve

Consider again the expression for the true stress-strain curve in the form

$$S=S_0(e\delta/n)^n.$$

It is evident that for a particular specimen with a given value of n, other than the limiting case of $n=0$, there is one value of strain for which the true stress is numerically equal to the tensile strength S_0, namely

$$\delta_0=n/e. \tag{20}$$

If a true stress-strain curve is drawn, it is interesting to locate this point and compare it with the true stress S_c at the point of instability, as shown in Fig. 3.

One possible use of this relation is to determine n when the tensile strength is known, or vice versa. In general, of course, the true stress-strain curve is plotted from data which include the force and strain at the point of instability. Since the relations derived in this report are based on the property that the strain at this point is numerically equal to n, all quantities would actually be known. A more fruitful use of the two points shown on the above figure would be to plot the curve, obtain the strain δ_0 as shown, then compute the ratio δ_c/δ_0. This ratio should be equal to $e=2.72$, so that its actual value is a measure of whether the use of the relations derived in this report is justified for the particular specimen. It is also a measure, to some extent, of whether the use of the n-power law is justified. However, this should be first ascertained by a plot of $\log S$ *vs.* $\log\delta$ to be certain of the range in which the law is valid.

IV. COMPARISON WITH EXPERIMENTAL DATA

Figure 1 shows the position of some experimental data points with respect to theoretical curves computed from Eq. (18) for plastic strains of 0.05, 0.10, and 0.20, respectively. The data were taken from various curves published in the literature, except for the data on Mn-Mo steel obtained in this laboratory. The principal sources were C. W. MacGregor, published in the Trans. A.I.M.E. **24**, 212 (1937), and E. L. Bartholomew, Jr., published in the Trans. A.S.M.E. **31**, 590 (1943).

It is not surprising that there is good agreement between the experimental points and theoretical curves, since only smooth stress-strain curves from the above references were chosen which appeared to follow the n power law. Very little data were available for values of n below 0.05 or above 0.40. Further, curves for specimens with very low values of n given in the literature could not be read with desired accuracy. These data were obtained from several items easily available to the

writer, and do not represent a survey of all published results.

The usefulness of the curves lies in their description, as one generalized family of curves, of the behavior of all materials, both brittle and ductile. The Mn-Mo steel had a tensile strength of 203,000 p.s.i. and a value of $n=0.07$, while the alpha-brass shown had corresponding values of 37,000 p.s.i. and $n=0.38$. Thus the behavior of many materials can be reduced to a common denominator.

V. CONCLUSIONS

The revised form of the n-power law given by Eq. (16), and the generalized family of curves given by Eq. (18), provide a new description of the useful properties of materials subjected to plastic deformation. This should aid in understanding and condensing the information presented in the literature for various materials. Further, deductions from the revised form of the law allow a check on experimental data as to the suitability of applying the n-power law to them.

A Technique for the Preparation of Grease or Solid Samples Dispersed in Grease-Like Media for Examination with the Electron Microscope

A. Y. Mottlau
Standard Oil Development Company, Linden, New Jersey
(Received April 22, 1949)

The nature of grease as a dispersion of solid particles in oil is briefly discussed, and the value of the electron microscope in studying this solid phase is pointed out. The requirements of a satisfactory grease specimen mounting technique are outlined and a new technique fulfilling these requirements is described. Electron micrographs are presented which demonstrate the usefulness of the technique in disclosing pertinent information on grease structure. The factors involved in the preparation of powder specimens for electron micrography are discussed in relation to the adoption of a modification of the grease mounting technique for this purpose. The modified technique for powders is outlined and illustrated with micrographs.

THE purpose of this paper is to describe the technique used at the Esso Laboratories for the electron microscope study of grease structure, and to describe a modification of this technique used for the past two years or more in the study of powdered materials such as catalysts and rubber fillers. A grease consists of an oil, thickened by a finely divided, dispersed solid. Generally this solid is a soap but greases are now being used in which the solid is completely inorganic. The determination of the size, shape, and state of aggregation of these solid particles or fibers can be adequately handled only with the help of an electron microscope. The problem has been found to be one of considerable importance, because the morphology of this solid material has a marked effect on the performance of the grease.

A satisfactory procedure for the examination of particulate solids with the electron microscope requires that the particles be dispersed on the specimen mount in such a manner that a minimum amount of overlap of one particle with another occurs. Conveniently, in a grease the solid is already in a satisfactory state of dispersion. It is only necessary that the dispersion be transferred to the specimen mount. In order to examine dry powders, however, the particles must first be dispersed. The simplicity and speed of the grease mounting technique to be outlined below suggested that a substance having the consistency of a grease would be an excellent medium in which to disperse dry powders, thereafter using the grease mounting technique to prepare the final electron microscope specimen. The details of this procedure will be discussed in the second half of this paper.

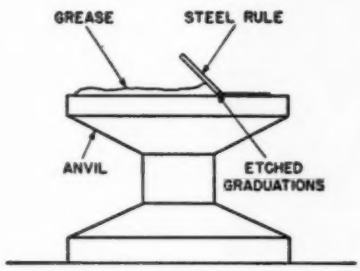


FIG. 1. Schematic diagram showing method of preparing striated grease film on top of anvil with a steel machinists' rule.

PREPARATION OF GREASE SPECIMENS

Requirements

A satisfactory technique for the examination of grease structure should meet the following requirements:

1. The grease should be deposited on the supporting membrane in such an amount and thickness that completely dispersed particles appear largely as discrete

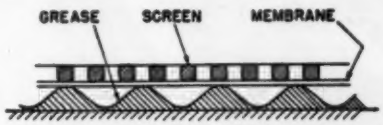


FIG. 2. Schematic diagram showing relative positions of specimen screen, membrane, and striated grease film during "printing" process.

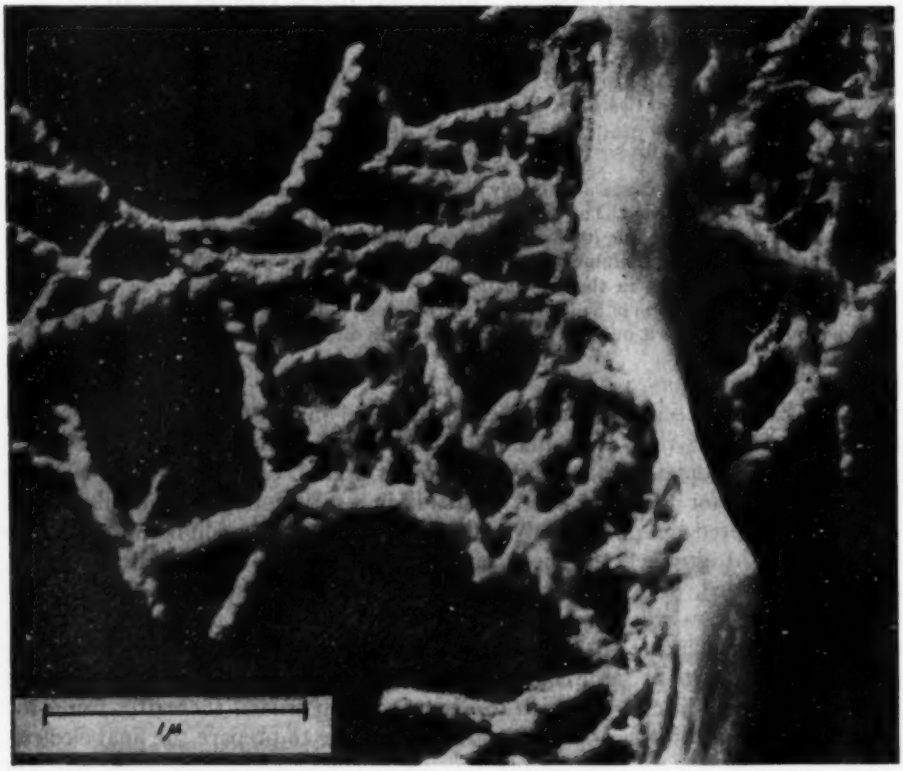
particles, and that undispersed agglomerates do not suffer any appreciable further dispersion or agglomeration.

2. The oil, not having any visible structure, should be removed so that it cannot impair the definition of the solid material.

3. The method should not involve contact with water because some greases contain water soluble soaps.

Two techniques for the preparation of grease specimens for electron microscopy have been described by Ellis.[1] One involves the deposition of the grease solids on the supporting membrane from ether suspension. The objection to this procedure was that it involved a step (the suspension of the grease solids in ether) that could readily affect the original degree of dispersion of those solids. In the other method described, the grease was spread in a thin film on a glass slide that had been previously coated with a resin. The thin resin film was then floated off on water. A modification of this procedure was used in this laboratory for some time in the study of water insoluble greases. It became necessary to abandon it, however, when interest turned to the examination of greases containing water soluble soaps.

In the procedure finally developed the grease is "printed" in thin, closely spaced lines directly on the supporting membrane which had been previously



(a)

(b) (c)

FIG. 3. Boiled lime soap grease showing the peculiar twisted soap fibers. In (a) and (b) where the specimen was shadowed with gold at 5:1, the fibers appear to consist of two fibers twisted as in a rope. In the unshadowed micrograph (c), the density pattern of the individual fibers does not confirm such a conclusion. Both pictures could be explained, however, by the assumption that the fibers are relatively thick "ribbons" twisted on their own axis. Compare with the obviously twisted "ribbons" in Fig. 4(a).

[1] S. G. Ellis, Can. J. Research 25, 119–23 (1942).

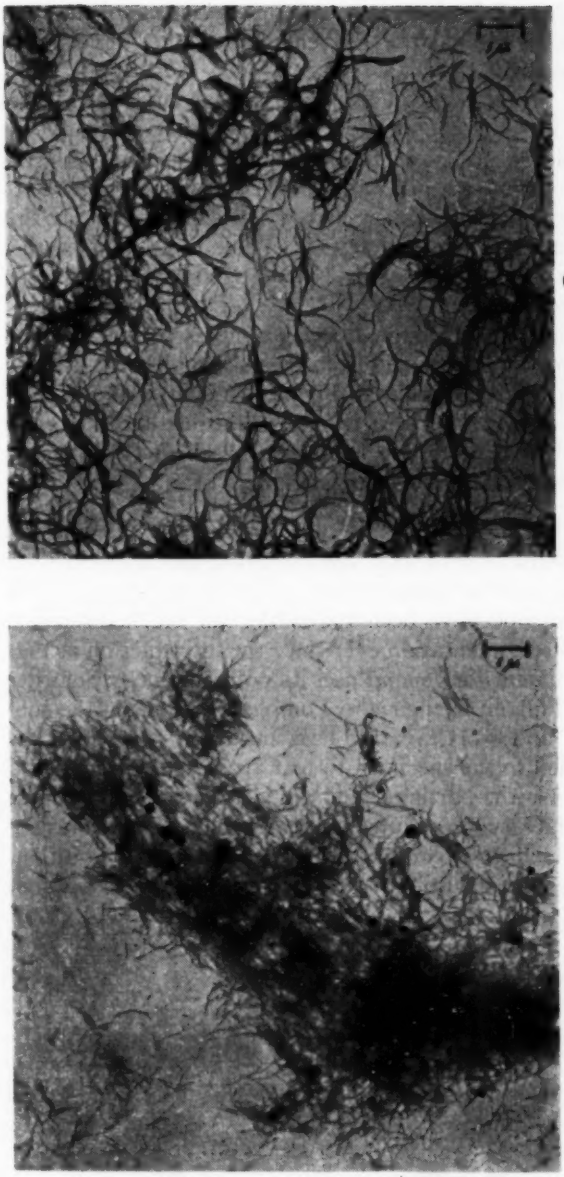


FIG. 4. (a) and (b) are sodium-barium greases of identical formula. The differences in structure are due to a difference in the cooling rate used in their manufacture. (a) Slow cooled. (b) Quick cooled.

mounted on a standard specimen screen. The method might be considered a refinement of the technique, due to Farrington and Birdsall,[2] by which the grease is applied directly to the screen mounted membrane with the aid of a thin glass rod.

Printing Technique

1. A small portion of the grease sample is spread on a flat polished surface such as a glass plate or the top of the chromium plated anvil sold with the R.C.A. electron microscope.

2. A thin, uniformly striated grease film is formed by sliding the tilted edge of a flexible steel machinist's rule across the flat surface of the anvil; see Fig. 1. If the grease is firm or "buttery" a rule having 100th inch etched markings is used. If the grease is soft (having a tendency to flow or settle) then a 64th inch rule will be found to give more satisfactory specimens.

3. In the next step, lines of grease are printed on a Formvar membrane which has been mounted on a specimen screen of the electrolytic type. This is done by placing the screen, membrane side down, on the striated grease surface, see Fig. 2, and tapping lightly with a needle. The membrane is made by casting on a glass slide from a 0.1 percent solution of Formvar in ethylene chloride, and subsequently floating it off on distilled water. The membrane must be mounted on an electrolytic type screen because the Formvar will adhere much more tenaciously to its relatively rough surface than to the shiny surface of the woven screen. Furthermore, care must be taken, if the screen is not absolutely flat, that the convex side be placed against the membrane, otherwise the striated grease film will not contact the center of the screen during the printing step.

4. The specimen screen is removed from the grease by bracing one side with a needle while thrusting another

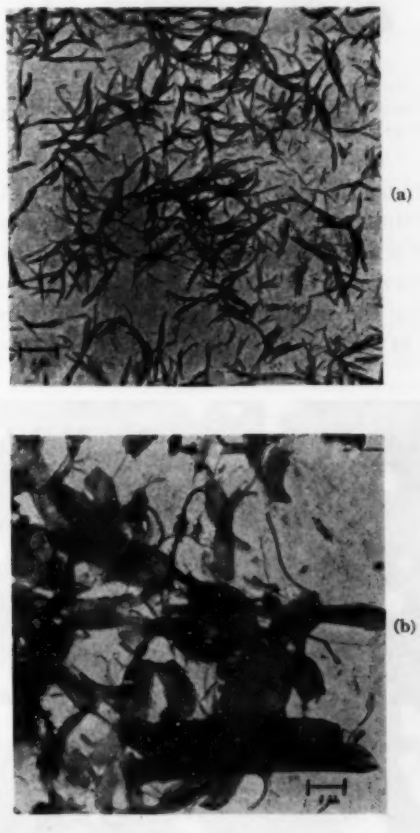


FIG. 5. (a) and (b) are sodium-barium greases of the same basic formula, differing only in the type of lubricating oil used in their manufacture.

[2] B. B. Farrington and D. H. Birdsall, Oil Gas J. 45, No. 46, 268–70, 275–9 (1947).

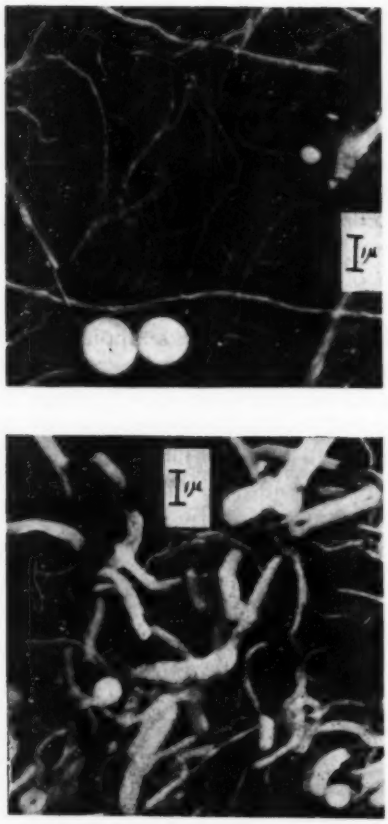


FIG. 6. Sodium greases gold shadowed at 5:1. The only difference between (a) and (b) was the molecular weight of the fatty acids used in their manufacture.

needle under the edge of the opposite side of the screen in a direction parallel to the stria. Held in a tilted position, the screen may then be handled with tweezers.

5. Removal of the oil from the grease is accomplished by placing the specimen screen for about three to five minutes in a Petri dish containing ASTM precipitation naphtha.* A microscope slide on the bottom of the Petri dish will faciliate the removal of the screen with tweezers.

6. Upon removal from the oil solvent, which dries quickly by evaporation, the specimen is ready for shadow-casting[3] or immediate examination with the electron microscope. It has been found good practice to examine the specimen with a light microscope at this point. With experience it is possible to predict whether or not the specimen will be satisfactory for electron microscope use.

Figures 3 through 6 are examples of the electron micrographs obtained from grease samples mounted by the printing technique. Figure 4(b) shows the type of agglomerate which, it is felt, would be altered or scattered by a mounting technique in which the grease is dispersed in a liquid such as ether.

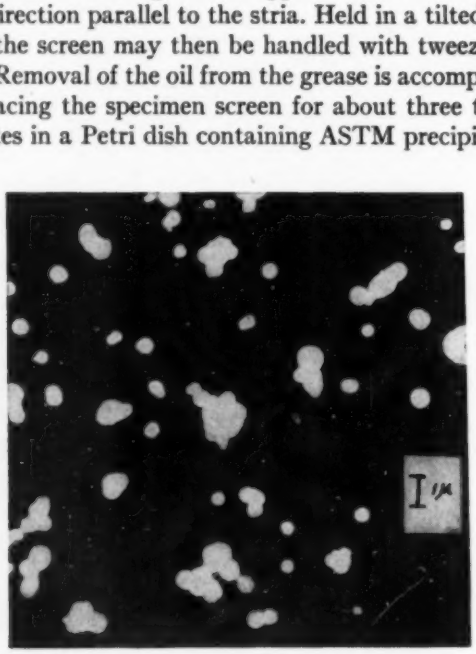


FIG. 7. Thermax carbon black, gold shadowed at 5:1.

PREPARATION OF POWDER SPECIMENS

Means of obtaining well-dispersed specimens of powders, such as pigments or catalysts, for the electron microscope have been the subject of much discussion and experimentation ever since the introduction of that instrument for general use. As pointed out by Schuster and Fullam[4] in their very complete treatise on the subject, and by Cravath,[5] it is necessary that work be applied to the specimen to tear apart agglomerates and disperse the particles. For this reason, one of the most satisfactory procedures currently used is to work the powder into a concentrated cellulose nitrate solution, subsequently forming a thin membrane of this dispersion.

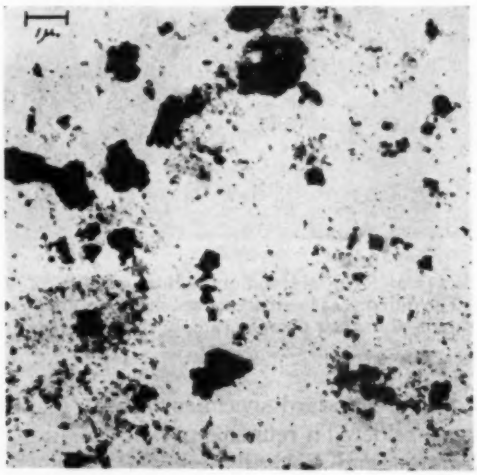


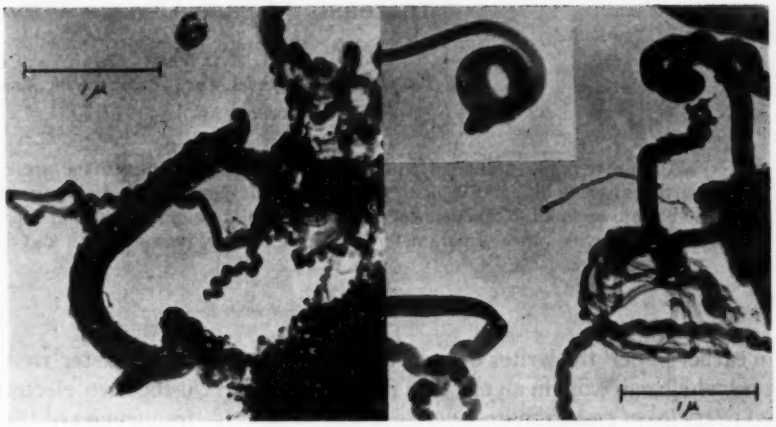
FIG. 8. This used catalyst represents a type of electron microscope specimen that is very difficult to prepare by the usual methods because of its wide spread in particle size and the low density carbonaceous material associated with it.

* Any clean, low boiling point, paraffinic naphtha can be used in this step.

[3] R. C. Williams and R. W. G. Wyckoff, J. App. Phys. **17**, 23 (1946).

[4] M. C. Schuster and E. F. Fullam, Ind. Eng. Chem. (Anal. Ed.) **18**, 653–7 (1946).

[5] A. M. Cravath, J. App. Phys. **17**, 1125–26 (1946).



(a) (b)

FIG. 9. Brown cuprene prepared by passing acetylene over a special iron catalyst. The printing technique has not marred the definition or the characteristic structure of these fibers. The catalyst granule in (a) shows small new fibers growing from its surface as pointed out by Watson (J. H. L. Watson J. App. Phys. **17**, 996–1005 (1946)).

Because the particles are imbedded in the supporting membrane, it is not possible to shadow-cast such a specimen. For the same reason the definition of the particles is somewhat impaired.

The observation that soap particles from conventional greases and carbon particles from the new carbon black greases as obtained by the printing technique, appeared to be well dispersed, suggested that the principles of this procedure might be applied to the preparation of powdered specimens. White petrolatum was found to be a convenient medium in which to disperse powdered samples. It had the further advantage of being completely soluble in ASTM precipitation naphtha.

Printing Technique for Powdered Samples

A small portion of white petrolatum about the size of a large pea is placed on a hard, flat surface, preferably with a dull finish such as finely ground glass. Using a steel spatula, the powder to be examined is worked into the petrolatum in an amount sufficient to form a stiff, but not a dry, paste. If the maximum of dispersion is desired, it is helpful to add a few drops of Paratac* to the mix. This high molecular weight polymer will greatly increase the viscosity of the paste, imparting high shear to the agglomerates. Like the petrolatum, Paratac is soluble in naphtha, and will therefore be removed during the final leaching step of the printing technique. After about 15 to 30 minutes of working, the paste is diluted with approximately 2 to 3 times the original amount of petrolatum. Additional working is required to form a uniform distribution of the dispersed powder. The amount of dilution depends on the particle size of the sample. A large particle size requires much less dilution than a small particle size. If the final grease-like mix is too stiff or too sticky for satisfactory "printing," it may be improved in consistency by the addition of a few drops of a light, white mineral oil.

From this point on the procedure is the same as that used in mounting a true grease, with the exception that the printed specimen must remain in the naphtha about 10 minutes to insure removal of all the petrolatum.

Figures 7 through 9 are electron micrographs of various specimens mounted by the printing technique. It will be noticed that the specimens are quite clean and free of any extraneous material which might impair the definition of the individual particles.

In conclusion it might be well to point out a feature of the printing technique which, though not premeditated, has turned out to be very helpful. Due to the manner in which the sample is printed on the supporting membrane there is a periodic change in concentration of particles deposited across the specimen screen. This characteristic makes the tolerable range of concentrations of particles in the grease or petrolatum quite broad, because some region of the specimen can generally always be found in which the particle concentration is quite satisfactory for the purpose desired.

* Paratac is the trade name for an oil solution of polybutene sold by the Enjay Company, Inc., 15 West 51st Street, New York 19, New York.

Increasing Space-Charge Waves

J. R. Pierce
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received April 8, 1949)

An earlier paper presented equations for increasing waves in the presence of two streams of charged particles having different velocities, and solved the equations assuming the velocity of one group of particles to be zero or small. Numerical solutions giving the rate of increase and the phase velocity of the increasing wave for a wide range of parameters, covering cases of ion oscillation and double-stream amplification, are presented here.

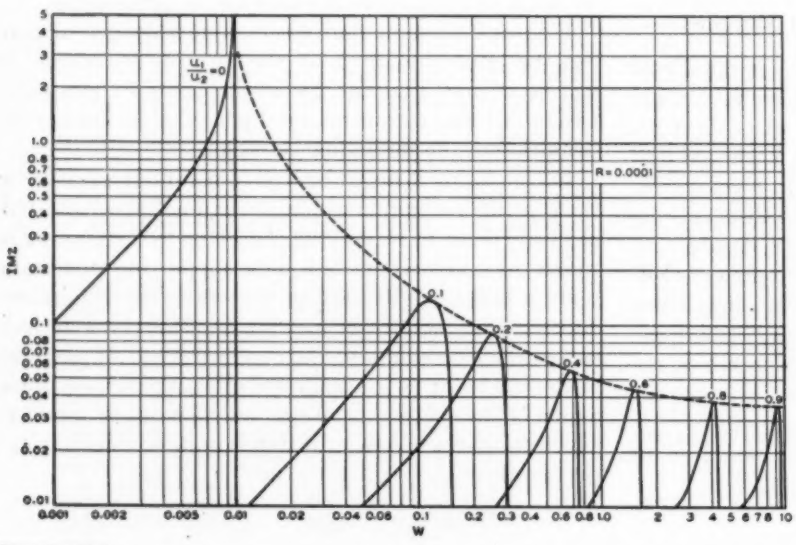
IN an earlier paper[1] the writer discussed the increasing waves which can exist in an electron flow made up of charged particles of two or more velocities. This original discussion was concerned with such waves as a cause of fluctuations in long electron streams. Subsequently, however, there has been a considerable interest in waves of this type from other points of view, both as a source of radio-frequency components of solar radiation,[2,3] and as a means for amplifying radiofrequency signals.[3–7] The equations given earlier[1] involve as parameters the velocities of the two or more components of space-charge and their plasma frequencies, but not the signs of the charges. Thus, these equations are the same as, or are simplified forms of, those given later.[3–7] The only difference is that a different range of parameters is involved in various problems. In the case of ion oscillations in electron beams, the electrons have a much higher plasma frequency than the ions and are moving much faster than the ions. In obtaining amplification using two electron streams, the velocities and plasma frequencies of the two streams can be nearly equal. It is not clear to the writer what range of parameters would be involved in the case of solar radiation.

Because of the increasing interest in increasing space-charge waves, it seems of some value to explore the nature of such waves for a wide range of parameters, embracing at least the range of ion noise and of double-stream radio frequency amplification, in order to compare behavior in various ranges. This is the purpose of the present paper.

We will start with (18) of reference 1.

$$1=\frac{\omega_e^2}{(\omega+j\Gamma u_0)^2}+\frac{\omega_i^2}{\rho_0}\int_{-\infty}^{u=\infty}\frac{d\rho_0}{(\omega+j\Gamma u)^2}. \qquad (1)$$

This equation applies to a one-dimensional or parallel-



FIGS. 1–6 and 14. The gain parameter, Imz, *vs.* the ratio W of the frequency to the plasma frequency of the faster stream for a given ratio R of the squares of the plasma frequencies of the slower and faster streams and for various ratios, u_1/u_2, of velocity of the slower stream to velocity of the faster stream. The wave amplitude varies with distance x as $\exp[Imz(\omega_{02}/u_2)x]$, where ω_{02} is the radian plasma frequency of the faster stream.

[1] J. R. Pierce, "Possible fluctuations in electron streams due to ions," J. App. Phys. **19**, 231–236 (1948).
[2] V. A. Bailey, Nature **16**, 599 (1948), J. Roy. Soc. NSW **82**, 107 (1948), Inst. J. Sci. Res. **A1**, 351–359 (1948), proposes a somewhat different type of increasing wave, that in an ion cloud in the presence of an electric field, as an explanation of solar radiation components.
[3] A. V. Haeff, "Space-charge wave amplification effects," Phys. Rev. **74**, 1532–1533 (1948).
[4] J. R. Pierce and W. B. Hebenstreit, "A new type of high-frequency amplifier," Bell Sys. Tech. J. **28**, 33–51 (1949).
[5] A. V. Hollenberg, "Experimental observation of amplification by interaction between two electron streams," Bell Sys. Tech. J. **28**, 52–58 (1949).
[6] A. V. Haeff, "The electron-wave tube—a novel method of generation and amplification of microwave energy," Proc. IRE (**7**, 4–10 (1949).
[7] L. S. Nergaard, "Analysis of a simple model of a two-beam growing-wave tube," RCA Review **9**, 585–601 (1948).

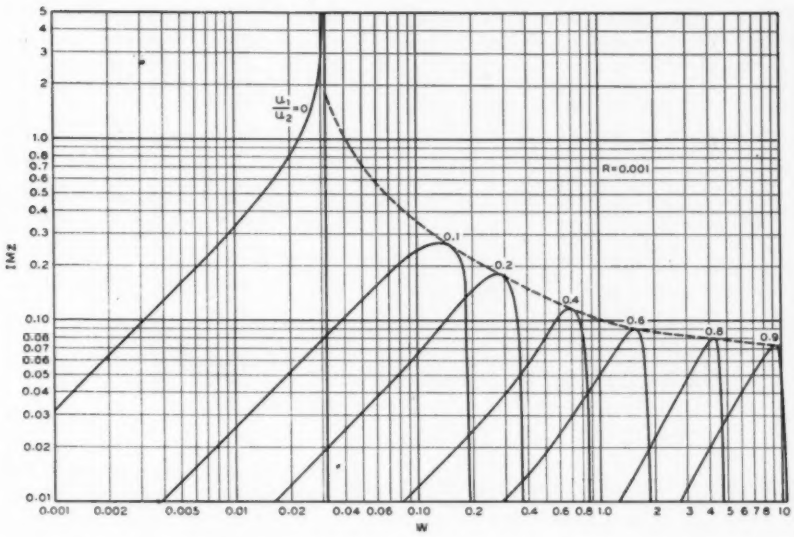


FIG. 2.

plane case, and applies to a small sinusoidal disturbance which is assumed to vary with time and distance in the x direction, the direction of flow, as $\exp(j\omega t-\Gamma x)$. Here ω_e is the electron plasma radian frequency and u_0 is the mean velocity of the electrons, while ω_i is the ion plasma radian frequency based on the total ionic charge density ρ_0, and $d\rho_0$ is an element of charge density associated with ions of velocity u.

We will consider a case of particles (electrons or ions) of two velocities, u_1 and u_2, only. Particles of these two velocities will have average charge densities $\pm\rho_{01}$ and $\pm\rho_{02}$, and charge-to-mass ratios $\pm(q/m)_1$, $\pm(q/m)_2$. Thus, the plasma radian frequencies ω_{01}^2 and ω_{02}^2 will be given by

$$\omega_{01}^2=\frac{(q/m)_1\rho_{01}}{\epsilon_0}, \tag{2}$$

$$\omega_{02}^2=\frac{(q/m)_2\rho_{02}}{\epsilon_0}. \tag{3}$$

Because Γ is nearly a pure imaginary, and in accord with the notation used in a later paper,[4] another variable β will be introduced

$$j\beta=\Gamma. \tag{4}$$

There are now particles of a velocity u_2 only involved in the integration in the last term of (1), and in the notation of (2)–(4), (1) becomes

$$1=\frac{\omega_{01}^2}{(\omega-\beta u_1)^2}+\frac{\omega_{02}^2}{(\omega-\beta u_2)^2}. \tag{5}$$

This equation will now be rewritten in terms of three

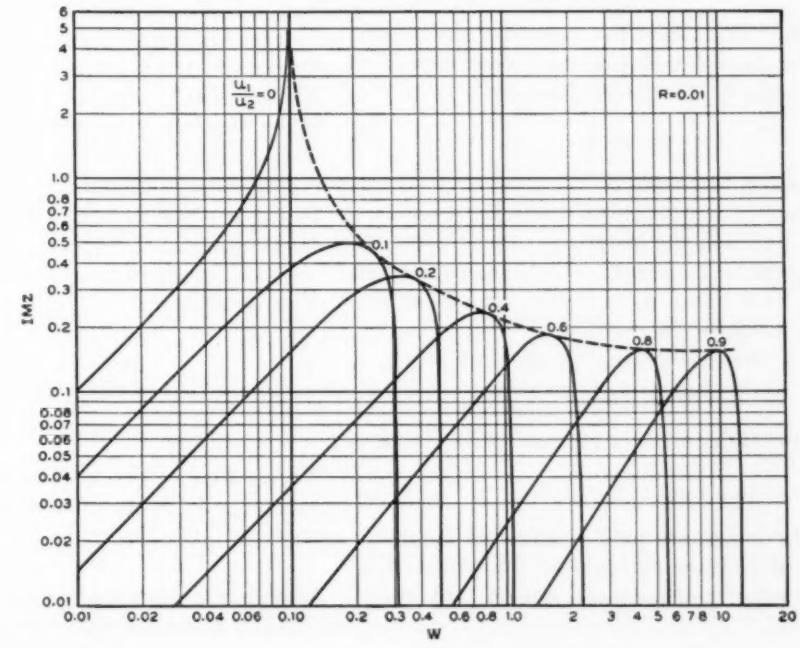


FIG. 3.

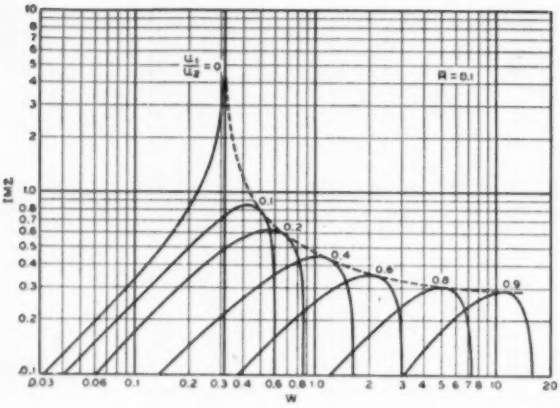


FIG. 4.

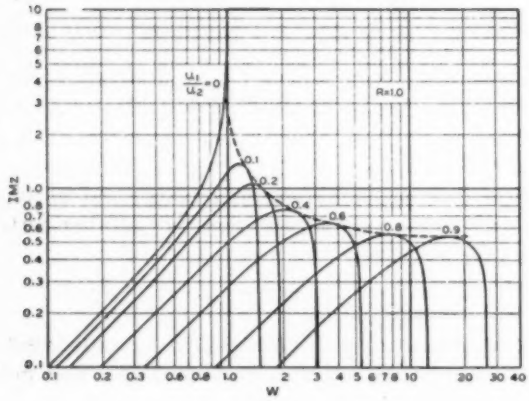


FIG. 5.

parameters chosen to have a simple physical significance in that they refer quantities to the plasma frequency and the velocity of the faster stream of particles.

$$R=(\omega_{01}/\omega_{02})^2. \tag{6}$$

Thus, R is merely the ratio of the squares of the plasma frequencies. For particles of the same charge-to-mass ratio, R is the ratio of the charge densities.

$$W=\omega/\omega_{02}. \tag{7}$$

Thus, W is merely the ratio of the frequency to the plasma frequency of one group of particles, which we shall choose to be the faster group.

The third parameter is the ratio of the mean velocities (u_1/u_2). As we have taken u_2 to be the velocity of the faster group of particles, $(u_1/u_2)<1$.

We will also choose a new variable z in place of β

$$z=\beta/(\omega_{02}/u_2). \tag{8}$$

The amplitude of the disturbance increases with distance as

$$\exp[Imz(\omega_{02}/u_2)x] \tag{9}$$

where Imz is the imaginary part of z (a real number). Thus, the rate of increase depends on the plasma frequency and the velocity of the faster stream, and on the parameter z.

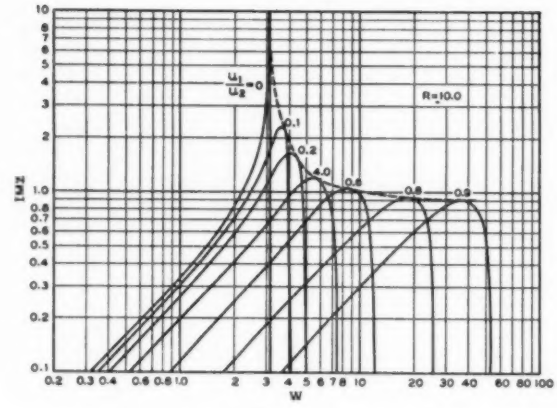


FIG. 6.

The real part of β is by definition the radian frequency divided by the phase velocity, v. Hence,

$$v=\left(\frac{W}{Rez}\right)u_2. \tag{10}$$

In terms of the new parameters and variable, (5) becomes

$$1=\frac{R}{(W-(u_1/u_2)z)^2}+\frac{1}{(W-z)^2}. \tag{11}$$

This equation has been solved numerically for a range of values of the parameters.

Let us consider first the behavior of gain with frequency. This is specified by the parameter Imz. The quantity $(\omega_{02}/u_2)x$ is a measure of distance defined in terms of the plasma frequency and the velocity of the faster stream and distance x. Thus, Imz is a measure of gain per unit distance, and if we take u_2/ω_{02} as a unit distance, the gain in this unit distance is $\exp(Imz)$.

In Figs. 1–6, Imz is plotted *vs.* W, the ratio of the frequency to the plasma frequency of the faster stream, for ratios of the squares of the plasma frequencies of the slower and faster streams, $R=0.0001$, 0.001, 0.1, 1, 10. The smallest of these values could represent the presence of gas ions in a charge density comparable to the charge density of a faster electron stream. Or, it could represent the presence of slower electrons forming a current of less than a ten-thousandth of the current carried by the fast electrons.

When the slower electrons are actually at rest, (11) assumes the form

$$1=\frac{R}{W^2}+\frac{1}{(W-z)^2}, \tag{12}$$

$$z=W\pm j(R/W^2-1)^{-\frac{1}{2}}$$

We see that at low frequencies the gain rises with frequency and goes to infinity at

$$W=(R)^{\frac{1}{3}}. \tag{13}$$

At higher frequencies z is real and there is no gain.

When the slow electrons are actually moving, the

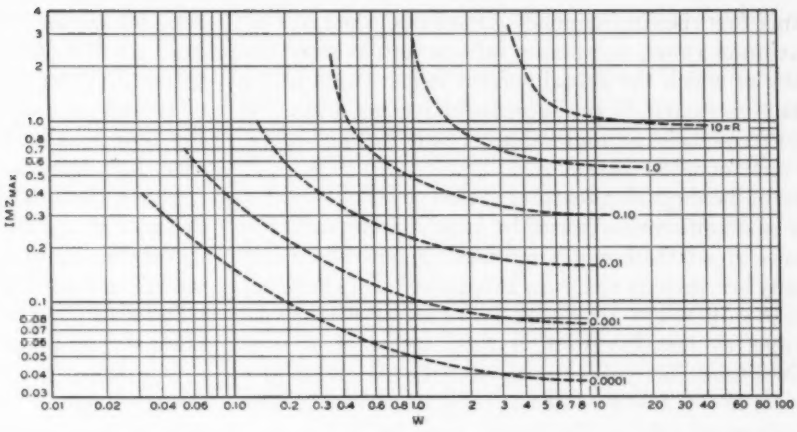


FIG. 7. The envelopes of the curves of Figs. 1–6, showing the largest values of Imz *vs.* W.

maximum gain is finite and occurs at frequencies higher than $(R)^{\frac{1}{2}}$, the frequency of maximum gain increasing as u_1/u_2 is made larger and the velocity difference becomes smaller. As the velocities approach equality the frequency of maximum gain approaches infinity and the maximum gain approaches a constant value which is a function of R.

For any velocity ratio u_1/u_2, at low frequencies W the gain per unit distance, (Imz), is proportional to frequency. At low frequencies the actual current is many times the minimum current necessary to produce double-stream gain, and the gain per wave-length should approach an asymptotic value, at least near $R=1$,[4] and hence the gain per unit distance should be proportional to frequency. Examining Fig. 5 ($R=1$) we see that this is so. Moreover, the maximum gain per wave-length (attained at low frequencies) should be proportional to the fractional velocity separation. We see from Fig. 5 that at low frequencies (say, $W=2$) going from $u_1/u_2=0.9$ to $u_1/u_2=0.8$, which about doubles the fractional velocity separation, doubles the gain also.

For a given value of R and u_1/u_2, the gain per unit

FIGS. 8–13 and 15. The ratio of wave phase velocity v to the velocity of the faster stream, u_2 *vs.* the ratio W of the frequency to the plasma frequency of the faster stream.

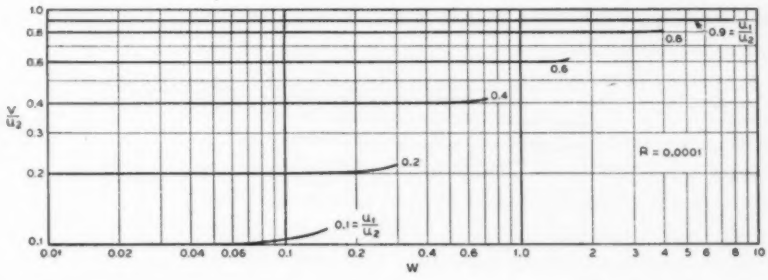


FIG. 9.

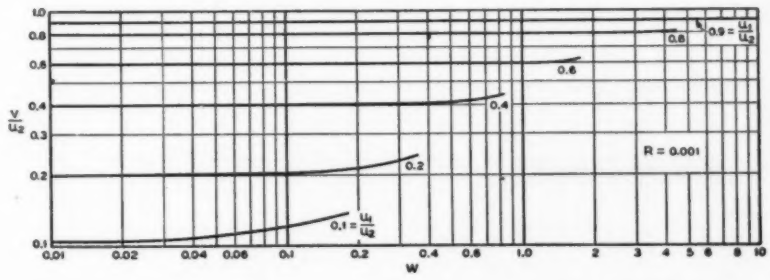
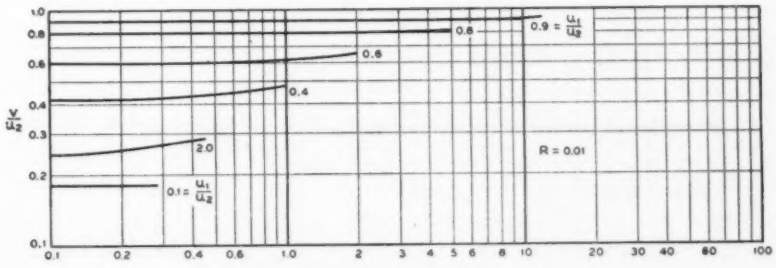


FIG. 10.



distance first rises as frequency is increased, then reaches a maximum value, and finally falls to zero at some frequency at which the actual current is just equal to the current necessary to produce an increasing wave. At higher frequencies all four values of z are real and represent unattenuated waves.

One of the surprising features of the curves of Figs. 1–6 is the small differences caused by large changes in R. For instance $R_1=0.0001$ might represent a ten-thousandth as many low velocity electrons as high-velocity electrons, while $R=10$ might represent 10 times as many slow electrons as fast electrons, a total range of R of a hundred thousand, yet for $u_1/u_2=0.9$ the maximum gain occurs at $W=9.2$ for $R=0.0001$ and at $W=37$ for $R=10$. The curve of gain *vs.* frequency (*Imz vs.* W) is considerably sharper for the very small values of R. This means that up to a point, greater fractional band widths will be attained in double stream amplifiers as the current in the slower stream is made larger. We see, however, that the curves for $R=10$ (Fig. 6) are sharper than the curves for $R=1$ (Fig. 5) and apparently the latter represents a region of optimum band width.

The envelopes of the curves of Figs. 1–6 are shown in Fig. 7. Here we see that at high frequencies (around $W=10$) *Imz* varies much less than R. For the highest frequencies shown, *Imz* changes by a factor of about 30 (34 db) for a change of R by a factor of 100,000.

From (9) we see that gain per meter is proportional to *Imz* times $\omega\gamma_2$. Consider a case in which W is around 10. First let $\omega_{01}/\omega_{02}=1$. Then let us consider a case in which $\omega_{02}/\omega_{01}=100$ ($R=0.0001$) and let ω_{01} be $\frac{1}{10}$ as large as before. In other words, ω_{02} has been changed by a factor of 10 and ω_{01} has been changed by a factor of $\frac{1}{10}$. From Fig. 3 we see that for the optimum value of u_1/u_2 in each case, *Imz* decreases by a factor of 0.065, while $\omega_{02}z$ decreases by a factor 10 times this, or 0.65. In this case, the decrease in ω_{02} by a factor of $\frac{1}{10}$ somewhat offsets the increase in ω_{01} by a factor 10, but the change in gain per meter in increasing the charge density of one stream by a factor of 100 and decreasing the charge density in the other stream by a factor of 1/100 is not very large.

One thing we see is that one can obtain "double-stream" gain with very small currents in the slower stream, and presumably with very small currents in the faster stream as well. Thus, one might obtain double stream gain because of primary emission from a negative control grid or secondary or primary emission from a positive control grid or other positive electrode, even in a tube intended to have only a single stream of electrons.

Also, we gather that ion oscillations could occur with a very small ion density.

Figures 8–13 show the ratio phase velocity to the velocity of the faster stream, v/u_2, as a function of frequency W for various ratios of the squares of the plasma frequencies, R, and for various values velocity ratios u_1/u_2. This ratio is given by (10) as (W/Rez).

We see from Fig. 8 that for the very low value of R, $R=0.0001$, the phase velocity is very nearly equal to the

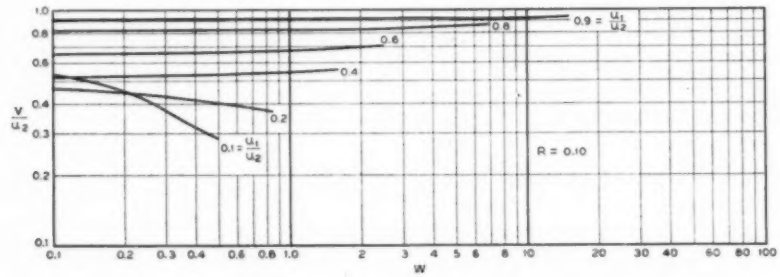


FIG. 11.

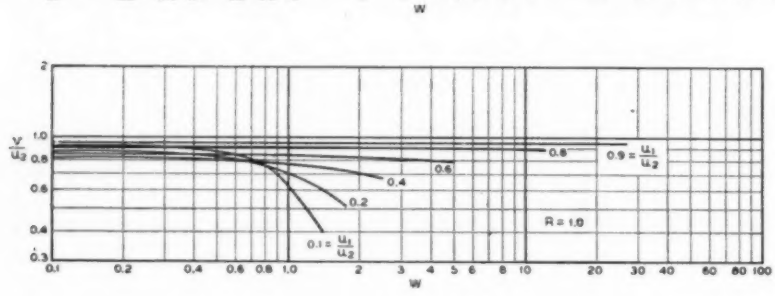


FIG. 12.

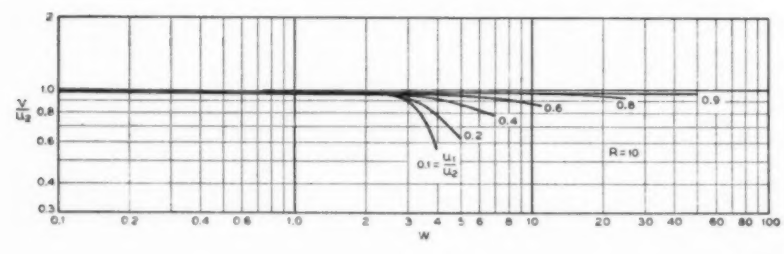


FIG. 13.

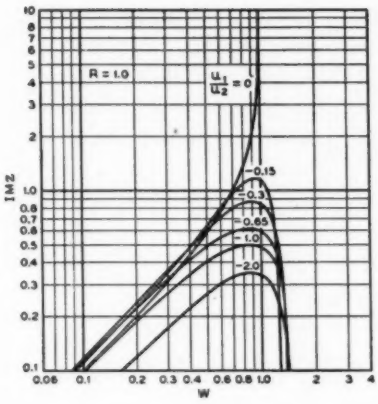


FIG. 14.

velocity of the slower electron stream for all velocity ratios from $u_1/u_2=0.1$ to $u_1/u_2=0.9$. There is a slight upturning as W approaches the limiting frequency at which the gain falls to zero.

While v/u_2 may initially fall as u_1/u_2 is decreased, (12) tells us that v/u_2 must eventually go to unity at $u_1/u_2=0$ and for $W<(R)^{\frac{1}{2}}$ (the region of increasing waves). This is illustrated in the curves for higher values of R. For instance, in Figs. 11 and 12, for the lower values of W v/u_2 finally rises as u_1/u_2 is decreased, after initially falling, and in Fig. 13 ($R=10$) we see that at the lower values of W, v/u_2 is nearly unity for all values of u_1/u_2.

The group velocity v_g is given in terms of the phase velocity v by

$$v_g=v\left(1-\frac{d(\ln v)}{d(\ln\omega)}\right)^{-1}. \tag{14}$$

On a logarithmic plot such as those of Figs. 8–13, the slope is unchanged by the scale factors, and hence $d(\ln v)/d(\ln\omega)$ is just the slope on these figures. On Figs. 8–10 ($R=0.0001$ to $R=0.01$) the slopes are all upward, so that the group velocities are a little larger than the phase velocities. The slope is never as great as unity, however, so the group velocity is always unity and the wave always travels to the right.

In Figs. 11–13 ($R=0.1$ to $R=10$), v/u_2 falls as W is increased, and sometimes rather steeply, giving a group velocity considerably less than the phase velocity. This is particularly true for smaller values of u_1/u_2.

The analysis of the double-stream amplifier for R nearly equal to unity[4] showed the phase velocity to be midway between u_1 and u_2. From Fig. 12 we see that for $R=1$, $u_1/u_2=0.9$, $v/u_2=0.95$, which is in accord with this.

One case is not covered in Figs. 1–13; that in which the two streams of particles travel in opposite directions. Curves for this case for $R=1$ (equal plasma frequencies) are given in Figs. 14 and 15. From Fig. 14 we see that the maximum gain occurs near $W=1$ (the plasma frequency of both streams in this case) and tends to shift to a little lower frequency and a lower value u_1/u_2 is made more negative.

In Fig. 15, let us first consider $-1<u_1/u_2<0$. For these velocity ratios, at low values of W the phase velocity is positive (in the direction of the faster electrons). It rises to infinity at a critical value of W which is a little less than unity, goes to minus infinity as W is increased and remains negative as W is further increased.

The magnitude of the phase velocity increases as u_1/u_2 approaches minus infinity. When u_1 is equal and opposite to u_2, the phase velocity is infinite: the "growing" disturbances merely change exponentially with distance.

At low values of W the group velocity is positive and nearly equal to the phase velocity. As W is increased, the group velocity goes through infinity to negative values where the slope on Fig. 15 is unity, passes through zero where the phase velocity goes to infinity, goes through infinity to negative values where the slope becomes unity again, and approaches the phase velocity, which is negative, for large values of W.

We see that for small values of W the energy flow is to the right. Then in a region a little below $W=1$ the energy flow is to the left. For another region above this the energy flow is to the right again. For still higher frequencies the energy flow is to the left.

We should remember that of two complex values of z which give waves which change with distance, one is the complex conjugate of the other. Thus, whichever way energy flows, there is an increasing wave if there is a complex root.

Figures 14 and 15 show curves for $u_1/u_2=-2$. In this case the faster electrons travel to the left, and the velocities shown in Fig. 15 have signs which in a given region are opposite those for the other curves. This is to be expected. If a wave travels in the direction of the faster electrons, for instance, and the direction of the faster electrons changes from right to left, the direction

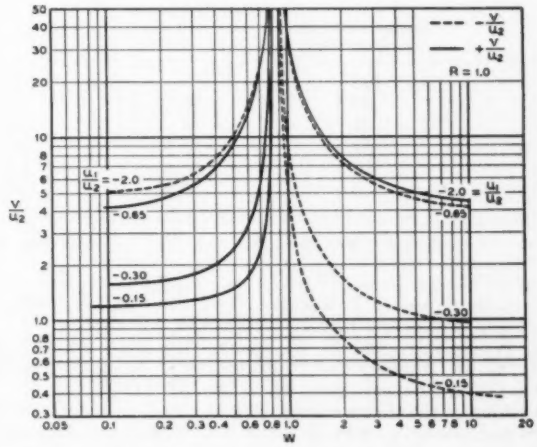


FIG. 15.

of travel of the wave changes. As a matter of fact, the curve for $u_1/u_2=-2$ could be replotted as a curve for $u_1/u_2=-0.5$.

It took a good deal of computing to obtain the curves of Figs. 1–15. It would be even more difficult to explore the more complicated case of a cylindrical beam with a magnetic focusing field, which was discussed in the earlier paper,[1] and no further information about the increasing waves associated with the plasma and cyclotron frequencies in this case have been obtained.

The writer wishes to acknowledge the contributions of Miss M. E. Moore in computing the roots of Eq. (11), and of Dr. L. R. Walker in devising a suitable method of computation.

The Effect of Cold Work in Metals on Powder Pattern Intensities

B. L. Averbach and B. E. Warren
Massachusetts Institute of Technology, Cambridge Massachusetts

(Received April 27, 1949)

Using crystal monochromated Cu $K\alpha$-radiation and a Geiger counter spectrometer, integrated intensity measurements have been made for powder samples of cold-worked and annealed α-brass. The effect of cold work is to increase the integrated intensity of the strong lines and make practically no change in the weak lines. The increase for the strong lines is presumably due to a reduction in extinction. Measurements were made of the (400) reflection on a cold-worked sample and on an annealed sample held at high temperature to demonstrate the difference between the effects of cold work and temperature vibration.

INTRODUCTION

CONSIDERABLE attention has been directed to the broadening of diffraction lines in x-ray patterns of cold-worked metals.[1–4] Two schools of thought have developed to explain the observed increase in line widths: the fragmentation theory ascribes the phenomenon to a reduction in particle size by cold work, and the microstress theory accounts for the same effect by a system of internal strains which are uniform over very small regions in the crystal, but which are random in magnitude, direction, and sign from one region to another. Additional information on the cold-working process can be obtained from a study of line intensities. Brindley and his collaborators,[3] using film techniques, compared the relative integrated intensities of cold-worked and annealed Cu, Ni, and Rh powders, and found that cold working decreased the line intensities except for the first line where there was an increase due to reduction in extinction. Since a progressively greater decrease was observed for the higher order lines, the effect was interpreted as a randomness in atomic position caused by cold work and has been described as a "frozen heat" randomness. The mean displacement from the lattice position was of the order of 0.08–0.10A. Since the interpretation of the effect of cold work on x-ray diffraction patterns is still in a vague and unsatisfactory state, it was decided to reinvestigate the effect of cold work on x-ray intensities with a Geiger-counter spectrometer and to remove cold work by careful annealing at closely spaced temperatures.

Table I. Stability of cold-worked brass intensity standards.

Line	Integrated intensity: Immediately after cold working	One year after cold working		Calculated
(111)	193	193	↔	193
(200)	89.5	91.5		93.6
(220)	48	51		50.5
(311)	57.5	55.5		57.1
(222)	18.3	21.8		16.5
(400)	9.6	9.7		10.0

Note.—↔ indicates where calculated and measured intensities were matched.

EQUIPMENT AND PROCEDURE

X-ray intensities on cold-worked and annealed brass (70 Cu, 30 Zn) were measured with a scaling circuit over 5- or 10-minute intervals at the peaks and 1 or 2 degree intervals in the background. Cu $K\alpha$-radiation monochromated with a rock salt crystal was used, and care was taken to maintain constant tube current and voltage. The incident and diffracted beams made equal

Table II. Effect of cold work and annealing on integrated intensities of x-ray diffraction lines. Brass (70 Cu, 30 Zn)—cold-worked at 25°C.

Line	Calculated intensities		Annealing temperature °C: 25	100	200	225	250	300
(111)	193	↔ 193	184	182	170	162	148	
(200)	93.6	89.5	87	85	86	84	76	
(220)	50.5	48	48	48	46	47	43.5	
(311)	57.1	57.5	55	55	54	55	53	
(222)	16.5	18.3	19.1	17.1	15.7	16.1	15.7	
(400)	10.0	9.6	8.4	8.7	10.6	8.7	9.6	

Note.—↔ indicates where calculated and measured intensities were matched.

[1] C. S. Smith and E. E. Stickley, Phys. Rev. **64**, 191 (1943).
[2] Stokes, Pascoe, and Lipson, Nature **151**, 137 (1943).
[3] G. W. Brindley and F. W. Spiers, Phil. Mag. **20**, 882, 893 (1935); G. W. Brindley and P. Ridley, Proc. Phys. Soc. **50**, 501 (1938); G. W. Brindley, Proc. Phys. Soc. **52**, 117 (1940).
[4] A review of the effects of cold work in metals on diffraction line widths and intensities may be found in C. S. Barrett, *Structure of Metals* (McGraw Hill Book Company, Inc., New York, 1943), pp. 363–380.

angles with the specimen, and intensities from different specimens could be compared directly by simply interchanging samples. Intensities were compared from day to day by means of a standard specimen whose line intensities did not change on aging. Cold-worked brass turned out to be a suitable standard, and Table I shows that this material was stable for at least a year at room temperature.

The spectrometer specimens consisted of filings taken from a 50 percent cold-rolled bar, sieved to −325 mesh, and compacted into slabs 1 in.×½ in.×⅛ in. at 60,000 p.s.i. Compacts were then annealed for one hour at selected temperatures up to 400°C under dry hydrogen, and back reflection photograms indicated that at all temperatures below and including 350°C the diffraction rings were uniform. Line spottiness was first detected at 400°C and such samples were not used in the spectrometer.

It was possible to establish the background intensity throughout the entire pattern, and the integrated intensity assigned to each line was all of the intensity above this background under each peak. The bases of the diffraction lines were considerably broader than had been anticipated and spread out over several degrees.

RESULTS

Figure 1 shows examples of the powder patterns from a cold-worked and annealed brass obtained by this method. In Table I the integrated intensities for two brass samples, one measured immediately after cold working and the other after aging for a year, are compared with calculated intensities. The calculations include a temperature correction,[5] a Lorentz-polarization factor corrected for a crystal monochromated beam,[6] and a dispersion correction,[7] and take no account of extinction. The calculated intensities in arbitrary units were matched to the observed values at only one line in one pattern, and all of the rest of the observed intensities were then compared directly. It is evident from Table I that cold-worked brass is a stable material, and that the observed intensities agree with the calculated values reasonably well.

In Table II the integrated intensities of the cold-worked and annealed specimens are compared. The most noticeable effect is that as the annealing temperature rises, the intensity of the strongest reflection (111) is appreciably reduced. The weaker reflections are correspondingly less reduced, and at a weak reflection such as (400), the intensities are approximately the same for the cold-worked and the annealed materials.

These results can be described in terms of primary extinction, and it is seen that annealing introduces extinction whereas heavy cold working seems to substantially eliminate it. The extinction coefficient, $\frac{I \text{ (with extinction)}}{I \text{ (without extinction)}}$ was calculated from the (111) diffraction line in each case and is plotted in Fig. 2. There is a sharp break in extinction at about 225°C and this coincides with a corresponding break in the line width and hardness values. These changes are usually associated with the formation of new unstrained grains (recrystallization), and the crystallite size calculated from the classical Darwin formula lies in the range $2\text{–}6\times10^{-5}$ cm.

The data shown in Table II were obtained with relatively narrow slits in order to accentuate the accuracy for the strong reflections. To obtain better data in

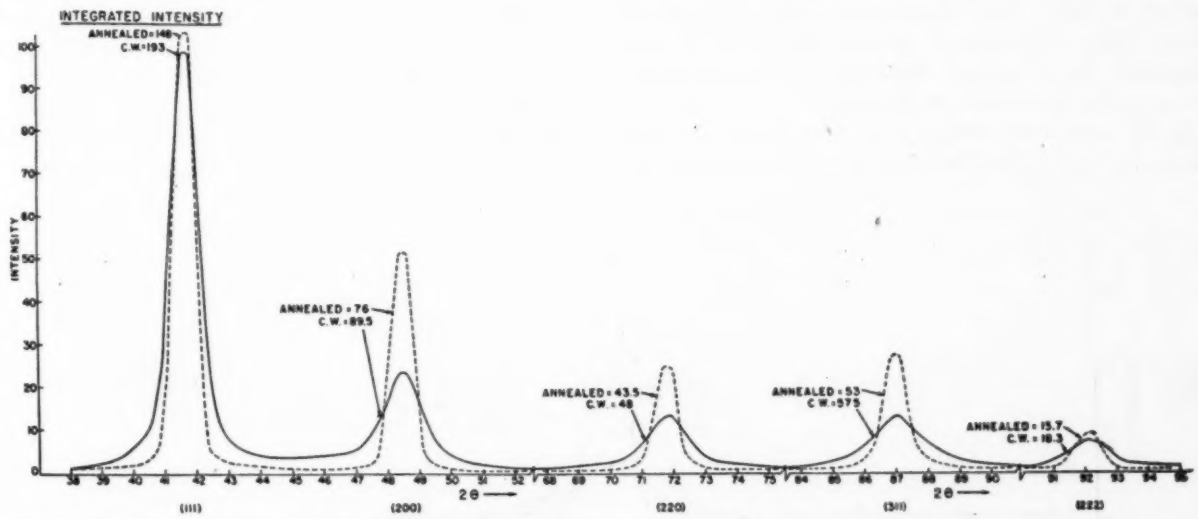


FIG. 1. Powder patterns of cold-worked and annealed brass. Effect of cold work on integrated intensity of x-ray diffraction lines. Brass (70 Cu–30 Zn). ——— cold-worked; - - - - - annealed at 300°C–1 hr.

[5] Debye-Waller correction, Debye temperature =313°K.

[6] $\frac{1+\cos^2 2\theta_c \cos^2 2\theta}{\sin 2\theta \sin\theta}$, θ_c = Bragg angle for monochromating crystal, θ = Bragg angle for sample.

[7] Cu, $\Delta f = 2.52$; Zn, $\Delta f = 2.16$.

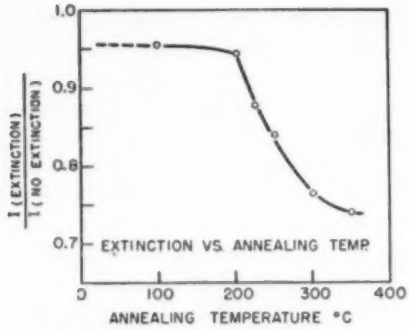


FIG. 2. Change of extinction coefficient with annealing temperature for cold-worked brass.

the back reflection region, the slits were opened wider and the (400) peak, as well as the background on each side, was carefully scanned. Figure 3 compares the (400) line of the cold-worked and annealed specimens (350°C). The pattern of the annealed specimen was quite sharp, and there was a considerable stretch of flat background on either side. The cold-worked specimen showed a broad peak and a continuously falling intensity for some eleven degrees (in 2θ) on the low angle side. On the high angle side interference with the next line, (331), prevented the true background from being reached, and the slowly rising tail beginning at about 119° was associated with this line, which is much more intense than the (400). If the steadily falling intensity adjoining the diffraction line is assigned to the line, and if the true background is established as the point where the cold worked and annealed background coincide, the integrated intensities for both lines, which are listed in Fig. 3, are identical within the experimental error. A background taken above this point would seem completely arbitrary, and almost any area could be obtained. From these data it is concluded that in weak reflections the integrated intensities from cold-worked and annealed specimens are identical.

As the annealing temperature was raised the long tails at the base of the (400) line seemed to contract, and the corresponding area was added at the peak. Table III shows, however, that the integrated intensities remained constant as the annealing temperature was raised. The half-maximum line width decreased sharply at 225°C, and this coincides with the sharp change observed for the extinction coefficient at the same temperature.

All of the previous patterns had been taken at 25°C. In order to compare the atomic displacements involved in cold work with those of heat motion, a pattern of the annealed specimen was taken at 350°C. The specimen was heated in a furnace with beryllium windows, an atmosphere of dry nitrogen was maintained, and the temperature was measured by a thermocouple in contact with the specimen. The 350°C pattern is also plotted in Fig. 3 and is shifted so that all of the peaks coincide. It is seen that temperature motion produced an entirely different pattern than did the cold-worked specimen. In the hot pattern the background is higher, but flat, and the line width is unchanged. A comparison of the line widths in Fig. 3 indicates that the patterns of the annealed specimens at 25 and 350°C have the same width while the cold-worked line is about three times as wide. If the integrated intensity of the hot pattern above the flat background is compared with that of the room temperature specimen, a ratio of 0.67 is obtained, compared to the value 0.71 expected from the Debye-Waller theory.

DISCUSSION

The atomic displacements involved in cold work are sufficiently different from those resulting from temperature vibration to produce measurably different effects on the powder pattern peaks. When intensity measurements are made far enough out on either side of the peak to accurately locate a background level, it is found that the peak areas for the large angle reflections are not decreased by cold work in contrast to the decrease which results from temperature vibration.

The effect of cold work on the integrated intensity of a powder pattern line is to increase the value for the strong reflections and produce practically no effect on

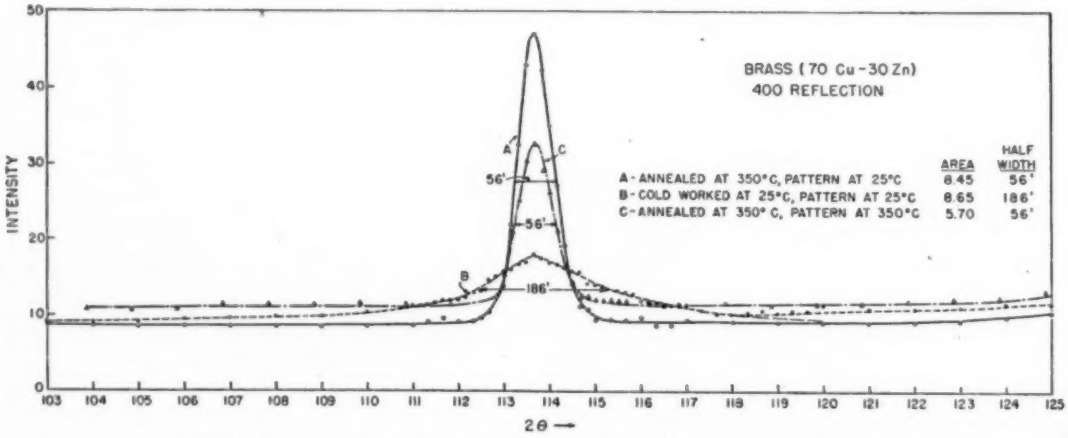


FIG. 3. Intensity near (400) reflection for cold-worked and annealed brass.

the weak reflections. This increase must be due to a reduction in primary extinction.

It has been shown[1,2] that the measured line broadening is fairly well correlated with λ, $\tan\theta$, and the directional variation of Young's modulus, if the broadening is assumed to be due to uniform strains in the individual crystals, the magnitude of the strains varying from one crystal to another. However, this cannot be the whole story; in crystals which are large enough to give rise to extinction effects, uniform strains will have no effect on the extinction. Reduction in extinction can come only from non-uniform strains, or reduction in the size of the regions in which a uniform strain exists.

A picture of cold work which will satisfactorily explain both the line broadening and the increase in intensity will probably involve uniform strains, non-uniform strains, and the effects of plastic deformation such as the reduction in size of the regions in which a uniform strain exists. It is possible that the two effects of cold work: line broadening and increase in integrated intensity result from different aspects of the cold-work deformation. For example, plastic deformation might decrease the size of the individual crystals enough to reduce primary extinction, but not enough to produce any appreciable particle size broadening. The strains left in these crystals, which will be different for the different crystals, might then account for the major part of the line broadening. In addition, non-uniform strains such as those involved in bending, and those representing the localized distortion in the immediate vicinity of dislocations, can contribute to both the reduction in extinction and the line broadening.

TABLE III. Effect of cold work and annealing on integrated intensities and line widths of (400) reflection brass (70 Cu, 30 Zn).

Annealing temperature °C	Integrated intensity	Half-maximum line width (minutes in 2θ)
	Pattern taken at 25°C	
25	8.65	186
100	8.50	184
150	8.20	134
200	8.60	124
225	8.20	73
250	8.25	58
300	8.30	56
350	8.45	56
	Pattern taken at 350°C	
350	5.70	56

The authors would like to acknowledge the assistance of Mr. Sergei Lorris in the experimental phases of this program.

The Structure of Porous Materials from Gas Penetration Rates

F. A. SCHWERTZ
Mellon Institute, Pittsburgh 13, Pennsylvania*
(Received April 29, 1949)

Gas penetration rates have been used to study the structure of three commercial grades of sintered Pyrex glass. Specific surface areas, "average pore diameters," pore population densities, average particle diameters, and particle population densities have been calculated on the basis of the data obtained.

IMPORTANT information about the average geometrical properties of a porous material may be obtained from a study of gas penetration rates. The information obtained includes estimated values of specific surface area, average pore diameter, pore population density, and, if the porous material consists of a compacted or sintered powder, average particle size and particle population density. The equations relating these various quantities to each other are derived in the first section of this paper. In the second section, these quantities are in turn related to the law of flow of gases through fine capillaries. In the third section numerical values of the quantities are obtained on the basis of experimentally determined rates of penetration of gases through a number of fritted Pyrex glass disks. The last section contains some suggestions for further experimental work.

1. GEOMETRICAL RELATIONS FOR POROUS BODIES

To derive the relations existing between the geometrical quantities mentioned above, it is useful to think of a porous body as composed of identical, parallel pores each of cross-sectional area, a, and length, L, numbering N per unit of area measured perpendicular to the pore axes. The *porosity*, ϵ, of the material, defined as the volume of voids per unit volume of porous body, is then given by

$$\epsilon=\frac{NAaL}{AL}=Na, \tag{1}$$

where A is the total cross-sectional area of the body. Similarly, the *specific surface area*, S, defined as the surface area per unit volume of *solid material* is given by

$$S=\frac{NalL}{AL(1-\epsilon)}=\frac{Nl}{(1-\epsilon)} \tag{2}$$

where l is the peripheral length of the cross-sectional area of the pores.

In order to write the equation describing the flow of gases in fine capillaries in a form independent of the cross-sectional shape, there is needed, in addition to the above mentioned quantities, the *hydraulic radius*. This quantity, designated by the symbol m, is defined as the cross-sectional area of the pore divided by its "wetted perimeter," that is,

$$m=\frac{a}{l}=\frac{NAaL}{ALNl}=\frac{\epsilon}{Nl}. \tag{3}$$

From Eq. (2) it follows that

$$m=\frac{\epsilon}{(1-\epsilon)S}. \tag{4}$$

This equation is of fundamental importance because, for actually porous bodies, it is taken as the *definition* of the mean hydraulic radius. For a circular capillary of diameter, D, $m=D/4$ so that the expression

$$D=\frac{4\epsilon}{(1-\epsilon)S} \tag{5}$$

may be used to calculate an average pore diameter which is equivalent to the mean hydraulic radius. From the fact that

$$\epsilon=Na=N\frac{\pi D^2}{4} \tag{6}$$

the pore population density, N, may then be calculated.

If the porous material is composed of a compacted or sintered powder, a close relationship exists between the specific surface area and the average particle size. If the particles are spherical in shape, the relationship is particularly simple since the ratio of the surface area of the average sphere to its volume must be equal to the surface area per unit volume of *solid material*, that is, to S. Hence

$$d=6/S, \tag{7}$$

where d is the average diameter of the particles.

The number of spherical particles, n, per unit volume of *porous material* must then be given by

$$n\frac{\pi d^3}{6}=(1-\epsilon)$$

or

$$n=\frac{6}{\pi d^3}(1-\epsilon). \tag{8}$$

In the gas flow experiments one measures essentially the specific surface area and hence it is convenient to

* Fellowship on Coke-Plant Physical Technology sustained by Koppers Company, Inc., Pittsburgh, Pennsylvania.

express the derived quantities in terms of S and ϵ. There results then

$$D=\frac{4\epsilon}{(1-\epsilon)S}, \tag{9}$$

$$N=\frac{(1-\epsilon)^2S^2}{4\pi\epsilon}, \tag{10}$$

$$d=6/S, \tag{11}$$

and

$$n=\frac{S^3(1-\epsilon)}{36\pi}. \tag{12}$$

Similar equations may be developed if a cylinder is a better model of the particle shape than a sphere.

2. THE LAW OF GAS FLOW

The equation which describes the rate of flow of gases through a fine circular capillary of diameter, D, and length, L, is

$$\dot{g}=\frac{\pi D^4P_m}{128\mu LRT}+\frac{\pi^2D^3}{16L(2\pi MRT)^{\frac{1}{2}}}\left(\frac{2}{f}-1\right). \tag{13}$$

In this equation $\dot{g}$ represents the specific flow rate in moles per unit time per unit pressure difference. P_m is the mean pressure of the flowing gas and T the temperature. μ and M are the viscosity and molecular weight of the gas, respectively, and f is a parameter which is interpreted as the fraction of gas molecules diffusely reflected from the capillary walls.[1]

The theoretical basis for this equation has been discussed in a number of places.[1,2] It is sufficient to say here that the first term in Eq. (13) is an expression of Poiseuille's law of viscous flow as it applies to a compressible fluid, whereas the second term is a measure of the so-called "slip-flow" and it becomes important where the mean free path of the gas molecules is of the order of the diameter of the capillary. In fact, the ratio, r, of the second to the first term of this equation is given by[3]

$$r=\frac{8\lambda_m}{D}\left(\frac{2}{f}-1\right),$$

where λ_m is the mean free path of the gas at the mean pressure in the capillary. It should be pointed out that, except for a multiplicative constant, the second term in Eq. (13) is of the same form as Knudsen's equation[4] for the diffusion of gases through circular capillaries which applies where $\lambda_m \gg D$.

[1] Sherwood, Brown, DiNardo, and Cheng, J. App. Phys. **17**, 802 (1946).
[2] E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938), p. 293.
[3] J. J. Wolken and F. A. Schwertz, J. Gen. Physiol. **32**, 153 (1948).
[4] M. Knudsen, Ann. d. Physik (4) **28**, 75 (1908).

It is possible to write Eq. (13) in a form which is independent of the cross-sectional shape of the capillaries by using the hydraulic radius in place of the diameter. There results then

$$\dot{g}=\frac{m^2aP_m}{k_0\mu LRT}+\frac{\pi ma}{L(2\pi MRT)^{\frac{1}{2}}}\left(\frac{2}{f}-1\right), \tag{14}$$

where $k_0=2$ for a circular capillary but assumes values lying roughly between 1.5 and 3.0 for other shapes. For a porous body containing N identical, parallel pores per unit area it follows that

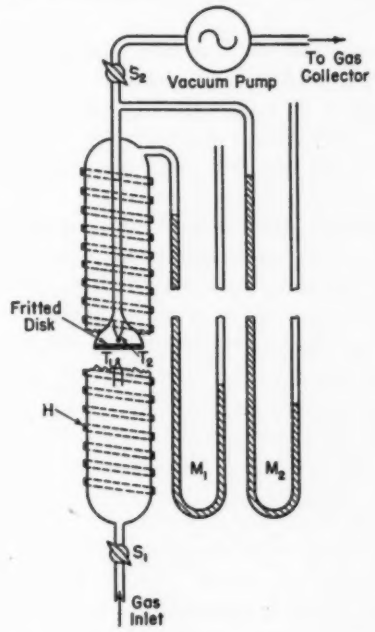


FIG. 1. Schematic drawing of experimental set-up.

$$NA\dot{g}=\frac{m^2aNAP_m}{k_0\mu LRT}+\frac{\pi maNA}{L(2\pi MRT)^{\frac{1}{2}}}\left(\frac{2}{f}-1\right).$$

With the aid of Eqs. (1) and (4) this may be written in the form

$$NA\dot{g}=\frac{\epsilon^3AP_m}{(1-\epsilon)^2S^2k_0\mu LRT}+\frac{\epsilon^2\pi A}{(1-\epsilon)SL(2\pi MRT)^{\frac{1}{2}}}\left(\frac{2}{f}-1\right).$$

As it stands, this equation cannot be applied to an actually porous body because the pores will neither be parallel nor equal in length to the porous body. Furthermore, the cross-sectional shapes of the pores will usually not be known. These uncertainties are taken into account by inserting constants k_v and k_d into this equation, so that finally

$$\dot{G}=\frac{\epsilon^3AP_m}{k_v(1-\epsilon)^2S^2\mu LRT}+\frac{\epsilon^2\pi A}{k_d(1-\epsilon)SL(2\pi MRT)^{\frac{1}{2}}}\left(\frac{2}{f}-1\right), \tag{15}$$

where $\dot{G}$ is the experimentally determined specific flow rate.

If flow experiments were carried out at pressures for which the mean free path of the flowing gas was small compared to the average pore diameter, only the first term in Eq. (15) would apply. In this case, the gas flow rate measured in volume units at the mean pressure and temperatures in the capillary would be given by

$$\dot{V}=\frac{\epsilon^3 A}{k_v(1-\epsilon)^2 S^2 \mu L}, \tag{16}$$

where $\dot{V}$ is the *specific* flow rate. This follows from the fact that $P_m\dot{V}=\dot{G}RT$. In the form in which it is written Eq. (16) applies equally well to gases or liquids. Hence experiments carried out on materials of known specific surface area and porosity yield information about the value of k_v.

P. C. Carman[5,6] has made a very careful experimental study of Eq. (16) using liquids on materials of known surface area. He found that a value $k_v=5.0$ led to values of the surface area which were accurate to within 3–4 percent for a wide variety of particle shapes.[6]

Where gases are used as the permeating fluid, Carman's value of k_v yields correct values of the specific surface areas for coarse powders. For fine powders, however, the values obtained with liquids are much larger than those obtained with gases. The reason for this situation was recognized independently by Rigden[7] and Arnell.[8] They corrected Eq. (16) with terms which are the equivalent of the second term in Eq. (15), and the results obtained for liquids and gases were brought into agreement. In the work which follows Eq. (15) is used with $k_v=5.0$. It is not necessary to know the value of k_d

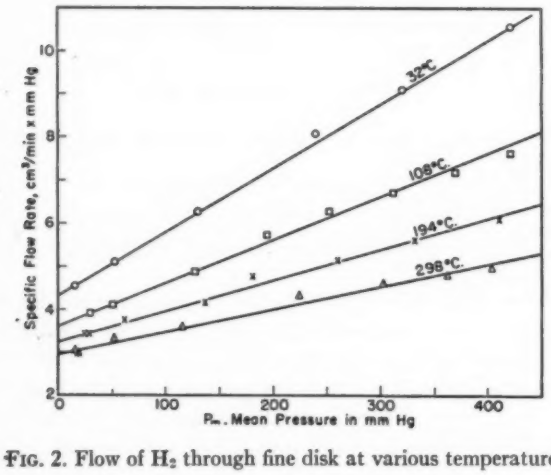


FIG. 2. Flow of H_2 through fine disk at various temperatures.

to calculate S, since if $\dot{G}$ is known for two values of P_m, S and k_d may be calculated simultaneously.

$\dot{G}$ is not a directly observable quantity for it is measured in moles per unit time per unit pressure difference. It is therefore useful for convenience of calculation to replace it by

$$\dot{G}=\frac{P_0\dot{F}}{RT_0},$$

where $\dot{F}$ is the specific flow rate in volume units corrected to standard conditions of temperature and pressure (0°C and 1 atmos.). If, in addition, one replaces $(1/k_d)(2/f-1)$ by Z, Eq. (15) becomes

$$\dot{F}=\frac{\epsilon^3 A T_0 P_m}{5(1-\epsilon)^2 S^2 \mu L T P_0}+\frac{\epsilon^2 \pi A Z R T_0}{(1-\epsilon)SLP_0(2\pi MRT)^{\frac{1}{2}}}. \tag{17}$$

Plotting $\dot{F}$ against the mean pressure, P_m, therefore yields a straight line the slope, α, of which is given by

$$\alpha=\frac{\epsilon A T_0}{5\mu L T P_0}\left\{\left(\frac{\epsilon}{1-\epsilon}\right)\frac{1}{S}\right\}^2 \tag{18}$$

and the intercept, β, of which is given by

$$\beta=\frac{\epsilon\pi A Z R T_0}{LP(2\pi MRT)^{\frac{1}{2}}}\left\{\left(\frac{\epsilon}{1-\epsilon}\right)\frac{1}{S}\right\}. \tag{19}$$

If Eqs. (18) and (19) are solved simultaneously, there results

$$\left(\frac{\epsilon}{1-\epsilon}\right)\frac{1}{S}=\left\{\frac{5\alpha\mu L T P_0}{\epsilon A T_0}\right\}^{\frac{1}{3}} \tag{20}$$

and

$$Z=\beta\left\{\frac{2MLP_0}{5\pi\alpha\mu\epsilon A R T_0}\right\}^{\frac{1}{2}}. \tag{21}$$

Since in the experiments to be described below, several gases were used at various temperatures, the calculational process has been carried one step further. α has been plotted against $1/\mu T$. Hence, if the dimensional changes with temperature are small, this procedure will also yield a straight line, the slope, γ, of which is given

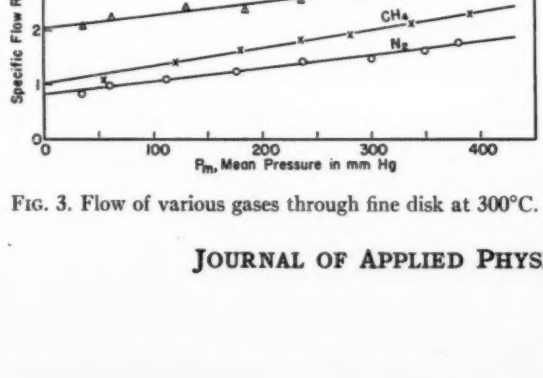


FIG. 3. Flow of various gases through fine disk at 300°C.

[5] P. C. Carman, Trans. I. C. E. London, **15**, 150 (1937).
[6] P. C. Carman, Trans. J. Soc. Chem. Ind. **57**, 225 (1938).
[7] P. J. Rigden, Nature **157**, 268 (1946).
[8] J. C. Arnell, Can. J. Research **24A**, 103 (1946).

by

$$\gamma=\frac{\epsilon AT_0}{5LP_0}\left\{\left(\frac{\epsilon}{1-\epsilon}\right)\frac{1}{S}\right\}^2. \tag{22}$$

Consequently

$$\left(\frac{\epsilon}{1-\epsilon}\right)\frac{1}{S}=\left\{\frac{5\gamma LP_0}{\epsilon AT_0}\right\}^{\frac{1}{2}}. \tag{23}$$

Also β has been plotted against $1/(MT)^{\frac{1}{2}}$. In this case, if Z is independent of the particular gas used, and of the temperature, the slope, δ, of the straight line so obtained is given by

$$\delta=\frac{\epsilon\pi AZRT_0}{LP_0(2\pi R)^{\frac{1}{2}}}\left\{\left(\frac{\epsilon}{1-\epsilon}\right)\frac{1}{S}\right\}. \tag{24}$$

Solving this equation simultaneously with (23) gives

$$Z=\delta\left\{\frac{2LP_0}{5\pi\gamma\epsilon ART_0}\right\}^{\frac{1}{2}}. \tag{25}$$

TABLE I. Slope and intercept values.

Gas	Experimental temperature, T °K	Slope, α cm³/sec. ×(mm Hg)²	Intercept, β cm³/sec. ×mm Hg
		Medium Porosity Disk	
H_2	301	94.7 $\times10^{-5}$	118 $\times10^{-3}$
H_2	377	57.0	113
H_2	463	39.8	100
H_2	578	30.0	88
He*	299	42.9	90
He	381	29.5	78
He	469	20.8	68
He	579	15.3	58
		Fine Porosity Disk	
H_2	305	21.5 $\times10^{-5}$	62.0$\times10^{-3}$
H_2	381	15.7	51.2
H_2	467	11.0	46.7
H_2	571	8.13	42.0
He	298	10.1	42.7
He	379	8.05	37.4
He	471	5.94	32.8
He	575	2.60	30.5
N_2	303	10.6	16.8
N_2	383	7.38	13.7
N_2	469	4.97	12.7
N_2	571	3.44	12.0
CH_4**	298	17.5	20.7
CH_4	377	10.9	19.0
CH_4	467	8.05	16.7
CH_4	577	5.32	15.0
		Ultrafine Porosity Disk	
H_2	300	1.67$\times10^{-5}$	14.0$\times10^{-3}$
H_2	385	1.29	12.3
H_2	489	0.90	11.2
H_2	576	0.81	10.1
He	303	0.65	11.1
He	376	0.67	9.4
He	487	0.44	8.7
He	576	0.29	7.9

* Helium 99.8 percent pure.
** Methane 99.0 percent pure.

3. EXPERIMENTAL RESULTS

Experiments were carried out on three different grades of fritted Pyrex glass designated by the maker (Corning Glass Works, Corning, New York) as *medium*, *fine*, and *ultrafine* "porosity." Each grade was in the form of a disk about one inch in diameter and about one-tenth of an inch thick, and each disk was the essential part of a so-called immersion tube. The tubes were mounted in the apparatus as shown in the sketch of the experimental set-up (Fig. 1). The pressure difference across the disk, as measured by the mercury manometers M_1 and M_2, was regulated by controlling the influx gas rate at the stopcock S_1. The gas which passed through the disk was continuously removed by the vacuum pump and, after dynamic equilibrium had been established, collected over water in a pneumatic trough. The temperature of the disk and flowing gas was controlled by varying the current in the heater coil H, and was measured by the thermocouples T_1 and T_2. The whole unit was thoroughly flushed out on changing from one test gas to another.

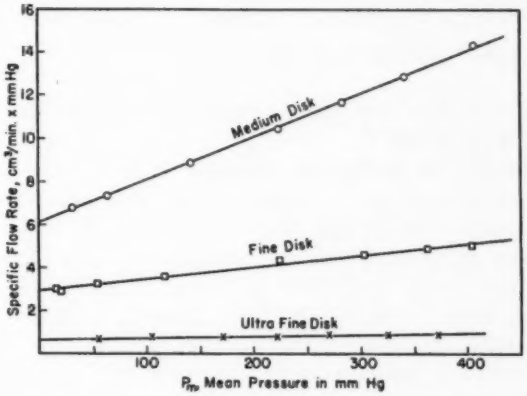


FIG. 4. Flow of H_2 through various disks at 300°C.

Typical of the results obtained are those shown in Fig. 2. Here the specific flow rate of hydrogen through the *fine* disk has been plotted in cm³/min. mm Hg. against the mean pressure, Pm. The gas volumes were measured at room temperature and atmospheric pressure. The data on this figure show that such a plot yields straight lines, the slopes and intercepts of which decrease with increasing temperature in accord with the predictions of the equations developed in the preceding section.

On Fig. 3 are given the experimental curves obtained on the *fine* disk at approximately 300°C for four test gases: hydrogen, helium, methane, and nitrogen. Again, as predicted by the theory, the intercepts of the straight lines decrease as the molecular weight of the test gas increases, and the slopes decrease as the viscosity of the gas increases.

In Fig. 4 is shown the variation in the specific flow rate for all three fritted glass disks. The test gas used

TABLE II. Geometrical constants for fritted Pyrex disks.

Disk	γ: $cm^3 \times poise \times °K$ / $sec. \times (mm\ Hg)^2$	δ: $cm^3 \times g^{\frac{1}{2}} \times °K^{\frac{1}{2}}$ / $sec. \times mm\ Hg \times mole^{\frac{1}{2}}$	ϵ*	A cm^2	L cm
Medium	231×10^{-7}	3.05	0.297	5.31	0.262
Fine	61.0×10^{-7}	1.43	0.267	6.55	0.285
Ultrafine	5.32×10^{-7}	0.342	0.253	4.83	0.145

* Obtained using a Hg densitometer.

TABLE III. Structure parameters for fritted Pyrex disks.

Disk	Specific surface area S, cm^2/cm^3	Slip factor Z	Average pore diameter D, microns	Pore population density N, millions / cm^3	Av. particle diameter d, microns	Particle pop. density n, millions / cm^3
Medium	2.11×10^3	0.615	8.0	0.59	28	58
Fine	3.57×10^3	0.550	4.1	2.0	19	300
Ultrafine	13.2×10^3	0.385	1.0	31.0	4.6	15,000

was hydrogen and the experimental temperature approximately 300°C.

Altogether thirty-two straight lines of the type presented on Figs. 2, 3, and 4, were obtained. Sixteen referred to the *fine* disk, eight to the *medium*, and eight to the *ultrafine*. Since, as mentioned above, the gas volumes were measured at room temperature and atmospheric pressure, and since these quantities varied slightly from one experiment to the other, the slopes and intercepts of the lines are not strictly comparable. These quantities were therefore corrected to normal conditions of temperature and pressure, that is, to 0°C and 760 mm of Hg. These corrected values have been collected in Table I. In addition to the temperature and pressure corrections, the time unit has been changed from minutes to seconds. The units applying to the slope values are cm^3 per sec. per $(mm\ Hg)^2$ and to the intercept values cm^3 per sec. per mm Hg.

The equations developed in the preceding section of this paper predict that a plot of α *vs.* $1/\mu T$ should yield a straight line provided the thermal expansion of the test disk is negligible. This provision is well satisfied for Pyrex glass. An α *vs.* $1/\mu T$ plot was accordingly made for each disk. The results are shown on Fig. 5. In each case a straight line passing through the origin is obtained. The slope, γ, of this line is related to the geometrical constants of the disk as brought out in Eq. (22).

Similarly a plot of β *vs.* $1/(MT)^{\frac{1}{2}}$ should produce a straight line provided Z is independent of the nature of the test gas and the temperature. The results in this case are illustrated in Fig. 6. Here again a straight line is obtained for each disk, the slope, δ, of which is related to the geometrical constants of the disk by Eq. (24).

The value of γ and δ for each disk are given in Table II along with values of the porosity, area, and thickness of each disk. This table, therefore, summarizes all the data required to calculate the specific surface area, S, and the "slip-factor" Z. The results are set forth in Table III. Also listed in this table are the derived quantities D, N, d and n which were calculated with the aid of Eqs. (9) to (12), inclusive, on the assumption that the glass particles were roughly spherical in shape. That this assumption is justified has been determined microscopically.[9] In addition, the porosity of each disk corresponds well with the figure for the closest packing of spheres (0.26).

An opportunity to check the accuracy of the above calculations is afforded by the data of Hodgins, Flood, and Dacey.[9] These experimenters used a sintered Pyrex glass plug fabricated from a powder for which the average particle diameter was known by microscopic observation to be 12.3 microns. Their data for hydrogen, helium and nitrogen are summarized in Figs. 7 and 8 for a plug having the following geometric constants:

$$\epsilon = 0.244,$$
$$A = 0.332\ cm^2,$$

and

$$L = 2.21\ cm.$$

From the straight lines on these figures γ and δ were found to have the values 1.02×10^{-8} and 4.5×10^{-3}, re-

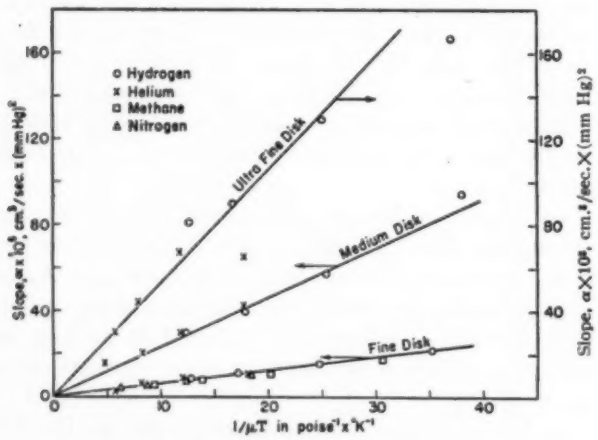


FIG. 5. Graphical representation of slope data.

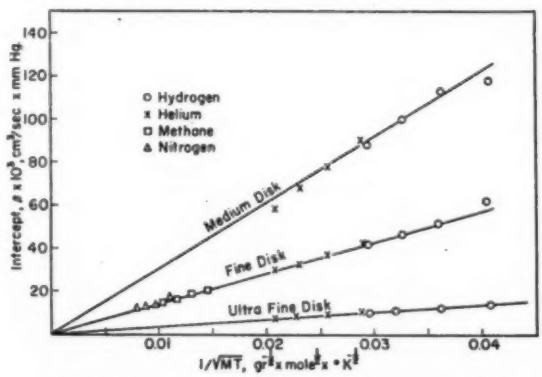


FIG. 6. Graphical representation of intercept data.

[9] Hodgins, Flood, and Dacey, Can. J. Research **24B**, 167 (1946).

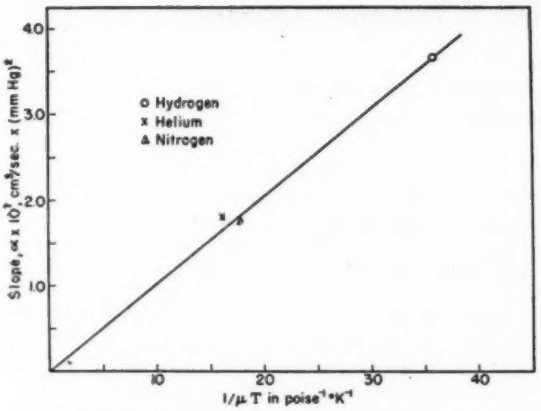


FIG. 7. Slope data for sintered Pyrex plug.

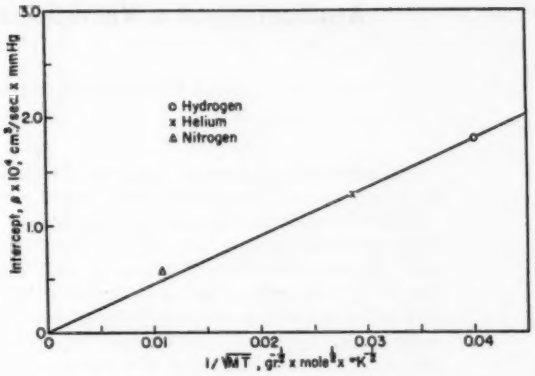


FIG. 8. Intercept data for sintered Pyrex plug.

spectively. These values and the geometric constants yielded the following structure parameters:

$$S = 6.00 \times 10^3 \text{ cm}^2/\text{cm}^3,$$
$$Z = 0.553,$$
$$D = 2.2 \text{ microns},$$
$$N = 6.7 \text{ millions/cm}^2,$$
$$d = 10 \text{ microns},$$

and

$$n = 1500 \text{ millions/cm}^3.$$

The value of $d = 10$ microns compares favorably with the microscopically determined value of 12.3.

4. DISCUSSION

Procedures similar to that described above have been employed by a number of investigators[5,10–14] to measure the specific surface areas of fine powders. In those cases where the areas were measured by an independent method, such as the Brunauer-Emmett-Teller adsorption technique[15] (B.E.T. method), they were found to be larger than those obtained by the gas penetration rate method. This implies that the gas measurements do not detect the fine pits and pores which may exist in the individual powder particles, but determine rather a *rounded-off* value of the mean particle size. In spite of this deficiency, the method is important because of the simplicity with which approximate values of specific surface areas may be obtained. Moreover, in a situation where many specific surface area determinations are to be made on the same material, it may be possible to correlate the results obtained from the penetration rate method with the more accurate but more time-consuming, B.E.T. method.

Up to the present time attempts to interpret the significance of the *slip-factor* Z in terms of the structure of porous materials have met with little or no success. On the basis of the above results, it appears that Z increases linearly with the particle size, d, and, therefore, linearly with the reciprocal of the specific surface area, S. In other words, the resistance to diffusive or "slip" flow increases with decreasing particle size.

The effect of porosity and particle shape on the *slip-factor* has not yet been investigated. It may well be that the value of Z is sensitive to the presence of fine pits and pores in individual particles. If so, Z may serve as a *correlation number* between the results obtained by the gas penetration method and the B.E.T. method. This possibility could be explored by making measurements on the same samples by both methods. The penetration rate measurements should be made at very high pressures where the flow is purely viscous and at very low pressures where the flow is purely diffusive, in order to determine k_v and k_d (or Z) independently.

[10] J. C. Arnell, Can. J. Research **25A**, 191 (1947).
[11] J. C. Arnell and G. O. Henneberry, Can. J. Research **26A**, 29 (1948).
[12] P. C. Carman and J. C. Arnell, Can. J. Research **26A**, 128 (1948).
[13] P. J. Rigden, J. Soc. Chem. Ind. **66**, 130 (1947).
[14] F. M. Lea and R. W. Nurse, Inst. Chem. Eng. London & Soc. Chem. Ind. Symposium on "Particle size analysis."
[15] Brunauer, Emmett, and Teller, J. Am. Chem. Soc. **60**, 309 (1938).

Application of a Variational Principle to Biconical Antennas*

C. T. Tai†
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received April 29, 1949)

A theoretical examination is made of the impedance of a biconical antenna, based upon a method devised by Schwinger in studying the discontinuity problems of a wave guide. In the present analysis, an integral equation to determine the aperture field is obtained by matching the tangential electric and magnetic fields along the boundary sphere. Using this integral equation, the effective terminating admittance of the antenna can be expressed in a form that is stationary with respect to small variations in the aperture field. It is shown that the zeroth-order solution of the admittance function thus obtained is the same as the one that Smith derived by neglecting all the high order waves in the interior region except the principal mode. For small-angle cones, the present formulation yields the same exact solution obtained previously, based upon several different methods. The paper also contains a discussion of the first-order approximate solution which is applicable to both small- and wide-angle cones. The analysis includes a detailed treatment of a method by which the characteristic values and characteristic functions for a given cone may be found approximately, but very accurately; as well as certain integrals involving the product of Legendre functions.

I. INTRODUCTION

SINCE the publication of Schelkunoff's work[1] in 1941, the problem of finding the input impedance of a biconical antenna has attracted considerable attention.[2–6] For small-angle cones, the problem appears to be very nearly closed.[4] No definite conclusion has yet been reached, however, for antennas of wide angle. In Smith's original work,[3] he obtained two approximate solutions for this latter case: the first takes into account the principal wave, one interior complementary wave, and two exterior complementary waves; the second** neglects all the interior waves except the principal wave but includes all the exterior complementary waves.

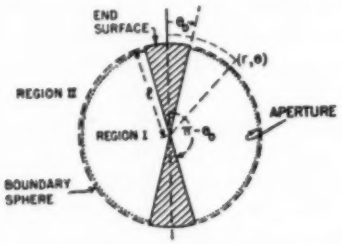


FIG. 1. The biconical antenna.

In the present analysis, examination of the problem is based upon a method originally devised by Schwinger in connection with the discontinuity problems of a wave guide. An integral equation to determine the aperture field is first obtained by matching the tangential components of the electromagnetic field along the boundary sphere. Using this integral equation, the function that defines the effective terminating admittance of the antenna can be expressed in a form which is stationary with respect to small variations of the aperture field. It is then shown that the zeroth-order solution to the problem is the same as Smith's second approximate solution. For small-angle cones, the "correction term" can be included that enables us to derive the same exact solution obtained earlier,[1,4] making use of several different methods. Finally, there is a discussion of the first-order solution showing how a better approximate solution can be obtained which is applicable to both small and wide-angle cones.

II. THE INTEGRAL EQUATION AND THE VARIATIONAL STATEMENT OF THE PROBLEM

The geometrical configuration of a biconical antenna is shown in Fig. 1. For convenience, the spherical surface defined by $r=l$ and $\theta_0 \leq \theta \leq \pi-\theta_0$ will be called the aperture of the system. The aperture is therefore part of the boundary sphere that is common to both Region I and Region II. As noted previously,[1–5] the E_θ- and H_ϕ-components of the electromagnetic field existing in Region I are given by

$$E_\theta = \frac{z_0 I_0}{2\pi r \sin\theta}[KY_t \sin\beta(l-r) - j\cos\beta(l-r)]$$

$$-\frac{jz_0}{2\pi r}\sum_n \frac{a_n}{n(n+1)}\frac{S_n'(\beta r)}{S_n(\beta l)}\frac{\partial L_n(\theta)}{\partial\theta}, \quad (1)$$

* The research reported in this paper was made possible through support extended Cruft Laboratory, Harvard University, jointly by the Navy Department, the Signal Corps of the U. S. Army, and the U. S. Air Force, under ONR Contract N5-ori-76, T. O. 1.

† Now at Stanford Research Institute, Stanford, California.

[1] S. A. Schelkunoff, "Theory of antennas of arbitrary size and shape," Proc. I.R.E. **29**, 493–521 (1941).

[2] S. A. Schelkunoff, "Principal and complementary waves in antennas," Proc. I.R.E. **34**, 23–32 (1946).

[3] P. D. P. Smith, "The conical dipole of wide angle," J. App. Phys. **19**, 11–23 (1948).

[4] C. T. Tai, "On the theory of biconical antennas," J. App. Phys. **19**, 1155–1160 (1948).

[5] C. T. Tai, "A study of the EMF method," J. App. Phys. **20**, 717–723 (1949); Technical Report No. 55, Cruft Laboratory, Harvard University (1948).

[6] C. H. Papas, "A theoretical investigation of spherically-capped antennas," doctoral thesis, Department of Engineering Sciences and Applied Physics, Harvard University, Cambridge (1948); "Input impedance of wide-angle conical antennas," Technical Report No. 52, Cruft Laboratory, Harvard University (1948).

** This solution was also obtained by S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York), p. 473.

$$H_\phi=\frac{I_0}{2\pi r\sin\theta}[\sin\beta(l-r)-jKY_t\cos\beta(l-r)]$$
$$-\frac{1}{2\pi r}\sum_n\frac{a_n}{n(n+1)}\frac{S_n(\beta r)}{S_n(\beta l)}\frac{\partial L_n(\theta)}{\partial\theta} \quad (2)$$

where

$$z_0=\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\doteq 120\pi \text{ ohms},$$

$$K=\frac{z_0}{\pi}\ln\cot\frac{\theta_0}{2}\text{ ohms},$$

$$S_n(\beta r)=(\beta r)^{\frac{1}{2}}J_{n+\frac{1}{2}}(\beta r),$$

$$S_n'(\beta r)=\frac{dS_n(\beta r)}{d(\beta r)},$$

and $L_n(\theta)$ belong to a class of Legendre functions which vanish at $\theta=\theta_0$, $\pi/2$, and $\pi-\theta_0$. The terms containing the trigonometric functions occurring in Eqs. (1) and (2) represent the electromagnetic field due to the principal wave propagating in Region I. The terms contained in the summation sign represent the complementary waves. The E_θ- and H_ϕ-components of the electromagnetic field existing in Region II are, likewise, given by

$$E_\theta=\frac{-jz_0}{2\pi r}\sum_k{}'\frac{b_k}{k(k+1)}\frac{R_k'(\beta r)}{R_k(\beta l)}\frac{\partial P_k(\theta)}{\partial\theta}, \quad (3)$$

$$H_\phi=\frac{-1}{2\pi r}\sum_k{}'\frac{b_k}{k(k+1)}\frac{R_k(\beta r)}{R_k(\beta l)}\frac{\partial P_k(\theta)}{\partial\theta} \quad (4)$$

where

$$R_k(\beta r)=(\beta r)^{\frac{1}{2}}H_{k+\frac{1}{2}}^{(2)}(\beta r),$$

$$R_k'(\beta r)=\frac{dR_k(\beta r)}{d(\beta r)}$$

and $P_k(\theta)$ denotes a Legendre polynomial of order k. The prime over the summation sign means that only odd values of k are considered.

As shown by Schelkunoff,[1] the problem of determining the input impedance of the biconical antenna is essentially to determine Y_t, the effective terminating admittance of the biconical transmission line formed by the lateral surfaces of the cones. The main difference between the present method and the previous ones is that instead of determining Y_t directly, we shall first derive an integral equation for the aperture field. Using this integral equation, we can express Y_t in a form that is stationary with respect to small variations of the aperture field. A detailed description of this method is found in Schwinger's original work,[7] or in lecture notes by Levine.[8]

[7] David Saxon, "Notes on lectures by J. Schwinger—discontinuities in wave guides," MIT Radiation Laboratory Report (1945).

[8] Harold Levine, "Notes on wave guide theory," Physics Department, Harvard University (1948).

Let the aperture field be denoted by $E_a(\theta)$. The continuity of the electric field across the boundary sphere requires that

$$\frac{-jz_0}{2\pi l}\sum_k{}'\frac{b_k}{k(k+1)}M_kP_k'(\theta)=\begin{cases}E_a(\theta), & \theta_0\le\theta\le\pi-\theta_0\\ 0, & 0\le\theta\le\theta_0 \text{ and}\\ & \pi-\theta_0\le\theta\le\pi\end{cases} \quad (5)$$

and

$$\frac{-jz_0I_0}{2\pi l\sin\theta}-\frac{jz_0}{2\pi l}\sum_n\frac{a_n}{n(n+1)}N_nL_n'(\theta)=E_a(\theta),\qquad \theta_0\le\theta\le\pi-\theta_0, \quad (6)$$

where the following abbreviations have been introduced:

$$P_k'(\theta)=\frac{\partial P_k(\theta)}{\partial\theta};\quad L_n'(\theta)=\frac{\partial L_n(\theta)}{\partial\theta};$$

$$M_k=\frac{R_k'(\beta l)}{R_k(\beta l)};\qquad N_n=\frac{S_n'(\beta l)}{S_n(\beta l)}.$$

Multiplying Eq. (5) by $P_r'(\theta)\sin\theta$ and integrating with respect to θ over the limits $\theta=0$ and π, we have

$$\frac{-jz_0}{2\pi l}b_rM_rI_{rr}=\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)P_r'(\theta)\sin\theta d\theta \quad (7)$$

or

$$b_r=\frac{j2\pi l}{z_0}\frac{1}{M_rI_{rr}}\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)P_r'(\theta)\sin\theta d\theta, \quad (8)$$

where

$$I_{rr}=\int_0^\pi P_r^2(\theta)\sin\theta d\theta=\frac{2}{2r+1},\quad r=1,3\cdots. \quad (9)$$

The direct integration of Eq. (6) gives

$$I_0=\frac{jl}{K}\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)d\theta. \quad (10)$$

Multiplying Eq. (6) by $L_m'(\theta)\sin\theta$, and then integrating, gives

$$a_m=\frac{j2\pi l}{z_0}\frac{1}{N_mI_{mm}}\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)L_m'(\theta)\sin\theta d\theta \quad (11)$$

where

$$I_{mm}=\int_{\theta_0}^{\pi-\theta_0}L_m^2(\theta)\sin\theta d\theta,\quad m=n_1,n_2,\cdots. \quad (12)$$

The continuity of the magnetic field across the aperture

requires that

$$\frac{-jKY_t I_0}{2\pi l \sin\theta}-\frac{1}{2\pi l}\sum_n \frac{a_n}{n(n+1)}L_n'(\theta)$$

$$=-\frac{1}{2\pi l}\sum_k \frac{b_k}{k(k+1)}P_k'(\theta),\quad \theta_0\leq\theta\leq\pi-\theta_0. \quad (13)$$

Substituting into (13) the value of b_r, I_0, and a_m given

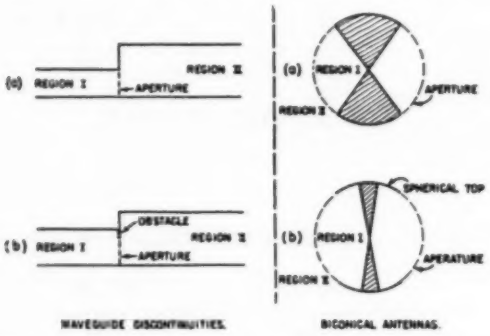


FIG. 2. Analogy between two discontinuity problems.

by (8), (10) and (11) we have

$$\frac{Y_t}{2\pi\sin\theta}\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)d\theta-\frac{j}{z_0}\sum_n\frac{L_n'(\theta)}{n(n+1)N_nI_{nn}}$$

$$\times\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)L_n'(\theta)\sin\theta d\theta$$

$$=\frac{-j}{z_0}\sum_k'\frac{P_k'(\theta)}{k(k+1)M_kI_{kk}}\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)P_k'(\theta)\sin\theta d\theta. \quad (14)$$

Equation (14) is the integral equation for the aperture field $E_a(\theta)$ that we have been seeking. Multiplying this equation by $\sin\theta E_a(\theta)$ and integrating over the aperture, and then dividing by $[\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)d\theta]^2$ we obtain finally the variational statement of the problem.

$$Y_t=\frac{j2\pi}{z_0}\frac{1}{\left[\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)d\theta\right]^2}\left\{\sum_n\frac{1}{n(n+1)N_nI_{nn}}\right.$$

$$\times\left[\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)L_n'(\theta)\sin\theta d\theta\right]^2$$

$$-\sum_k'\frac{1}{k(k+1)M_kI_{kk}}$$

$$\left.\times\left[\int_{\theta_0}^{\pi-\theta_0}E_a(\theta)P_k'(\theta)\sin\theta d\theta\right]^2\right\}. \quad (15)$$

The analogy between the expression of Y_t as derived here and the corresponding expression of B as given by Eq. (16), p. 126, in Schwinger's notes[7] for the wave guide problem is obvious. If we consider the physical nature of these two problems [Fig. 2(a)], the only difference is that for the wave guide problem, both Region I and Region II support a TEM-wave while for the biconical antenna this is true only of Region I. The present formulation can be easily extended to a spherically top-loaded biconical antenna, where the spherical cap plays the same role as an obstacle in a wave guide [Fig. 2(b)].

III. EVALUATION OF Y_t

To evaluate the value of Y_t as expressed by Eq. (15)' we can expand the aperture field in a complete set of orthogonal functions. Although the choice of this set of functions is quite arbitrary, nevertheless the accuracy of the result obtained by neglecting the high order terms depends upon their proper selection. Suppose we consider the following set,

$$E_a(\theta)=\frac{-A_0}{\sin\theta}+\sum_n A_nL_n'(\theta), \quad (16)$$

where $L_n(\theta)$ is the general Legendre function introduced before. Except for the undetermined coefficients A_n, the expression is of the same form as Eq. (1). Since the variational expression of Y_t is independent of the scale of the aperture field we can put $A_0=1$. Substituting Eq. (16) into Eq. (15), we obtain

$$Y_t=\frac{jz_0}{\pi K^2}\left\{-\sum_k'\frac{2P_k^2(\mu_0)}{k(k+1)I_{kk}M_k}+\sum_n\frac{n(n+1)I_{nn}A_n^2}{2N_n}\right.$$

$$-\sum_k'\sum_n\frac{2P_k(\mu_0)I_{nk}A_n}{I_{kk}M_k}$$

$$\left.-\sum_k'\sum_n\sum_m\frac{k(k+1)I_{nk}I_{mk}A_nA_m}{2I_{kk}M_k}\right\} \quad (17)$$

where

$$\mu_0=\cos\theta_0$$

and

$$I_{nk}=\int_{\theta_0}^{\pi-\theta_0}P_k(\theta)L_n(\theta)\sin\theta d\theta. \quad (18)$$

Denoting

$$Y_{t0}=\frac{-jz_0}{\pi K^2}\sum_k'\frac{2P_k^2(\mu_0)}{k(k+1)I_{kk}M_k}$$

$$=\frac{-jz_0}{\pi K^2}\sum_k'\frac{(2k+1)P_k^2(\mu_0)}{k(k+1)}\cdot\frac{R_k(\beta l)}{R_k'(\beta l)}, \quad (19)$$

$$\beta_n=\frac{jz_0}{2\pi K^2}\frac{n(n+1)I_{nn}}{N_n}, \quad (20)$$

$$\alpha_n=\frac{-jz_0}{\pi K^2}\sum_k'\frac{P_k(\mu_0)I_{nk}}{I_{kk}M_k}, \quad (21)$$

$$\gamma_{nm}=\gamma_{mn}=\frac{-jz_0}{2\pi K^2}\sum_k{}'\frac{k(k+1)I_{nk}I_{mk}}{I_{kk}M_k}, \tag{22}$$

we can write

$$Y_t=Y_{t0}+2\sum_n \alpha_nA_n+\sum_n \beta_nA_n{}^2+\sum_n\sum_m \gamma_{nm}A_nA_m. \tag{23}$$

Since Y_t is stationary with respect to the variation of $E_a(\theta)$, we determine the unknown coefficients A_n by putting $\partial Y_t/\partial A_n=0$, which gives

$$\alpha_n+\beta_nA_n+\sum_m \gamma_{nm}A_m=0, \quad n=n_1, n_2, \cdots. \tag{24}$$

This is an infinite set of linear equations that A_n has to satisfy. If we multiply Eq. (24) by A_n and sum over n,

$$\sum_n \alpha_nA_n+\sum_n \beta_nA_n{}^2+\sum_n\sum_m \gamma_{nm}A_nA_m=0. \tag{25}$$

Combining Eqs. (23) and (25), we finally obtain

$$Y_t=Y_{t0}+\sum_n \alpha_nA_n. \tag{26}$$

Mathematically, the term Y_{t0} can be treated as the zeroth-order solution to the problem. It is the same one that Smith obtained by neglecting all interior waves except the TEM mode. The sum of the remaining terms can be considered as the "correction term," with the coefficient A_n satisfying Eq. (24). Physically, we can interpret Y_{t0} as an admittance contributed by the interaction between the dominant wave and all the exterior complementary waves, and α_nA_n as an admittance contributed by the interaction between the nth-order interior wave and all the exterior waves. For small-angle cones, as we shall show later, only the coupling between waves of nearly the same order is significant.

IV. SMALL-ANGLE CONES

For small-angle cones, by putting $P_k(\mu_0)=1$, the zeroth-order solution is given by

$$Y_{t0}=\frac{-jz_0}{\pi K^2}\sum_k{}'\frac{2k+1}{k(k+1)}\frac{1}{M_k}. \tag{27}$$

It will be recalled,[1,4] however, that the exact solution for Y_t previously obtained is given by

$$Y_t=\frac{-jz_0}{\pi K^2}\sum_k{}'\frac{2k+1}{k(k+1)}\frac{1}{(M_k-N_k)}$$

$$=\frac{z_0\beta l}{2\pi K^2}\sum_k{}'\frac{2k+1}{k(k+1)}J_{k+\frac{1}{2}}(\beta l)H_{k+\frac{1}{2}}{}^{(2)}(\beta l). \tag{28}$$

Thus the so-called "correction term," which has a net effect of multiplying the zeroth-order solution by a factor of M_k/M_k-N_k to each individual term of the series, is by no means negligible. In fact, it can be easily verified that as $\mu_0\to 1$, (by making use of the limiting value of $I_{nk}I_{mk}/I_{nn}$ as discussed previously[4]) the system of equations defined by Eq. (24) can be solved. Thus,

$$A_n=A_r=\frac{1}{r(r+1)}\left(\frac{M_rN_r}{M_r-N_r}\right)\sum_k{}'\frac{2k+1}{M_k}\cdot\frac{I_{nk}}{I_{nn}}, \quad r=1, 3\cdots. \tag{29}$$

The "correction term" is then given by

$$\sum_n \alpha_nA_n=\frac{-jz_0}{\pi K^2}\sum_k{}'\frac{2k+1}{k(k+1)}\frac{N_k}{M_k(M_k-N_k)}. \tag{30}$$

Adding Eqs. (27) and (30), we obtain precisely the same expression of Y_t as that given by Eq. (28). The present formulation therefore yields the same exact result as derived earlier by several different methods. Now suppose we keep only one term of A_n in Eq. (24), and hence also one "correction term" in Eq. (26), then for small-angle cones the first-order approximate solution is

$$Y_{t1}=Y_{t0}+\alpha_1A_1=\frac{-jz_0}{\pi K^2}\sum_k{}'\frac{2k+1}{k(k+1)}\frac{1}{M_k}$$

$$-\frac{jz_0}{\pi K^2}\frac{3}{2}\frac{N_1}{M_1(M_1-N_1)}. \tag{31}$$

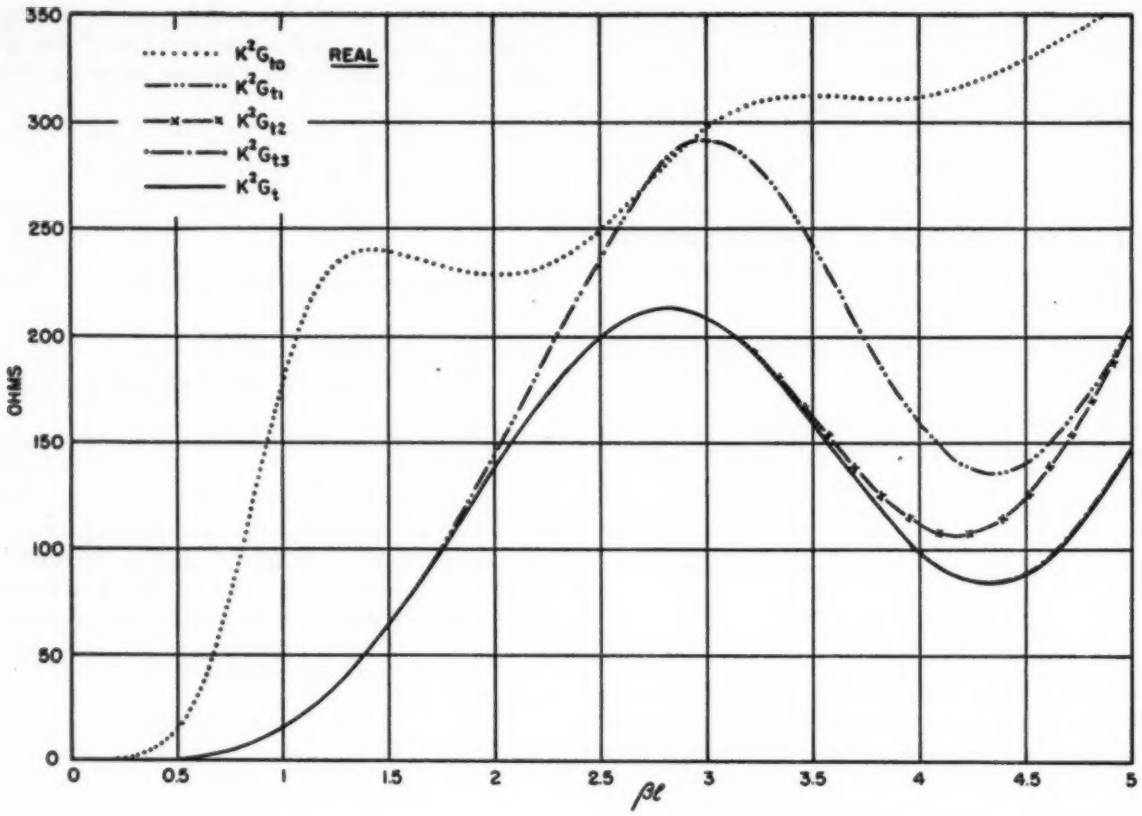
There will be no difficulty in obtaining the higher order approximate solutions by including more interior waves. Thus, the rth-order approximate solution takes the form

$$Y_{tr}=\frac{-jz_0}{\pi K^2}\sum_k{}'\frac{2k+1}{k(k+1)}\frac{1}{M_k}$$

$$-\frac{jz_0}{\pi K^2}\sum_{k=1}^{r}{}'\frac{2k+1}{k(k+1)}\frac{N_k}{M_k(M_k-N_k)}. \tag{32}$$

Values of K^2Y_t and K^2Y_{tr} up to $r=3$ for $\beta l\leq 5$ are plotted in Fig. 3. Since the series contained in Eq. (28) can be represented by a finite number of terms of sine and cosine integrals,[1,9] we have used the latter identity to evaluate K^2Y_t. For the series contained in K^2Y_{t0}, the sum is taken up to $k=39$, which gives an accuracy within about one percent. It is obvious from Fig. 3 that for small-angle cones, the zeroth-order solution differs considerably from the exact solution. The first-order solution, which is obtained by including one interior complementary wave, gives a much better approximation. It is also worth mentioning here that while the present approximate solutions, in general, approach the exact solution from the top, the earlier approximate solutions, i.e., $(K^2Y_t)_r$ of reference 4, approach it from

[9] S. O. Rice, "Sums of series of the form $\sum_0^\infty a_nJ_{n+\alpha}(z)J_{n+\beta}(z)$," Phil. Mag. 35, Series 7, 686–693 (1944).



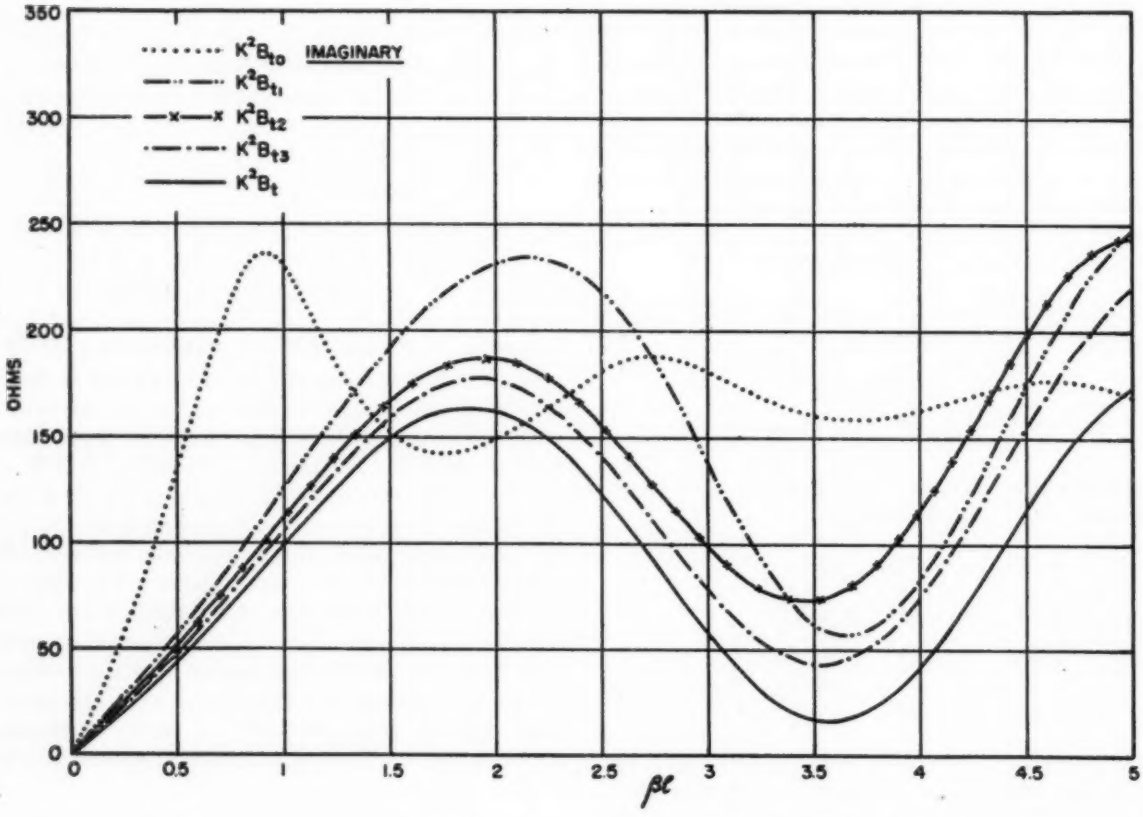


FIG. 3. A comparison of solutions of different order for small-angle biconical antennas.

below. The two approaches merge together as the number of interior waves considered is increased.

V. WIDE-ANGLE CONES

The foregoing technique to obtain the various approximate solutions can also be applied to wide-angle cones. In view of the fact that there is, so far, no exact solution for this case, the accuracy of the various approximate solutions can be judged only from the rate of convergence of the successive approximate solutions. Because of the practical difficulty in evaluating the fractional-order Bessel functions, numerical computation is further limited to the first two approximate solutions. Thus, the zeroth-order solution Y_{t0} is given by Eq. (19). The first-order solution is given by

$$Y_{t1}=Y_{t0}-\frac{\alpha_{n_1}{}^2}{(\beta_{n_1}+\gamma_{n_1n_1})} \tag{33}$$

where

$$\alpha_{n_1}=\frac{-jz_0}{\pi K^2}\sum_k{}'\frac{(2k+1)P_k(\mu_0)I_{n_1k}}{2M_k}, \tag{34}$$

$$\beta_{n_1}=\frac{jz_0}{\pi K^2}\frac{n_1(n_1+1)I_{n_1n_1}}{2N_{n_1}}, \tag{35}$$

$$\gamma_{n_1n_1}=\frac{-jz_0}{\pi K^2}\sum_k{}'\frac{k(k+1)(2k+1)I_{n_1k}{}^2}{4M_k}, \tag{36}$$

n_1 denotes the lowest characteristic value of n. To simplify the numerical computation, we consider only those values of θ_0 of which the corresponding values of n_1 are odd integers; then

$$I_{n_1k}=\frac{2n_1P_k(\mu_0)P_{n_1-1}(\mu_0)}{(k^2+k-n_1{}^2-n_1)}, \quad k\neq n_1 \tag{37}***$$

and

$$I_{n_1n_1}=\frac{4}{2n_1+1}\left[\frac{P_{n_1-1}(\mu_0)P_{n_1-2}(\mu_0)}{(2n_1-1)(2n_1-3)}\right.$$
$$\left.+\frac{P_{n_1-2}(\mu_0)P_{n_1-3}(\mu_0)}{(2n_1-3)(2n_1-5)}+\cdots+\frac{P_1(\mu_0)P_0(\mu_0)}{3\cdot1}\right]. \tag{38}***$$

For $n_1=3$, 5, 7, the corresponding values of θ_0 are respectively 39.23°, 57.43°, and 66.06°. Values of K^2Y_{t0} and K^2Y_{t1} for these three cones are plotted in Fig. 4, whereas in evaluating α_{n1} and $\gamma_{n_1n_1}$, we carry the sum up to $k=39$. Comparing Y_{t0} and Y_{t1}, we observe that the differences between them are, on the whole, not very large within the region $\beta l\leq 5$. For practical purposes, it is therefore quite satisfactory to use Y_{t1} as a measure of the terminating admittance for these three cones.

*** When n_1 is an odd integer, $L_n(\mu)$ reduces to $P_n(\mu)$. Equations (37) and (38) then are available in literature (See, for example, E. W. Hobson, *Spherical and Elliptical Harmonics* (Cambridge University Press, London, 1931), pp. 35–39. (Notice that the formula for I_{nn} as given in Hobson's has a misprint.)

Although we have so far limited our numerical computation to three particular cones, it is not difficult to extend it to cones of several other dimensions by a proper selection of the values of n. The mathematics involved is quite different, however. In the first place, for example, formulas like Eqs. (37) and (38) are no longer applicable to I_{nk} and I_{nn} if n is different from an odd integer. As the analysis involves some interesting aspects of the Legendre functions, the subject merits fairly comprehensive treatment here.

If we examine Eqs. (33)–(36), it will be seen that the essential functions involved in the computation are N_{n_1}, I_{n_1k}, and $I_{n_1n_1}$. The remaining functions $P_k(\mu_0)$ and M_k are well known, and have been tabulated. The function N_{n_1} which is the logarithmic derivative of $x^{\frac{1}{2}}J_{n_1+\frac{1}{2}}(x)$ is known numerically for $n_1>1$ only when n_1 is an integer, or a half-integer. Pending the tabulation of $J_{n_1+\frac{1}{2}}$ for an arbitrarily assigned value of n_1 our evaluation of Y_{t1} is then limited to $n_1=\frac{3}{2}$, 2, 5/2, 3, $\cdots$ etc. For this set of values of n_1, the simplest cases occur when n_1 is an odd integer; then $L_{n_1}(\mu)$ reduces to $P_{n_1}(\mu)$ and the integrals I_{n_1k} and $I_{n_1n_1}$ become very simple as we have shown in the foregoing section. For the rest of the values of n_1, our problem is a matter of evaluating these integrals. Of course it is also important to know the value of θ_0 or μ_0 for these values of n_1. In his work, Smith[8] interpolated the value of μ_0 for a given value of n_1 between the values obtained by plotting the roots in n for $P_n(\mu_0)$ and $Q_n(\mu_0)$. It should be remarked that the Legendre function of the second kind is the same as $L_n(\mu_0)$ when n is an even integer. Smith also worked out a series expansion for $L_n(\mu_0)$ in terms of the Legendre polynomials. By using a finite number of terms of the series, he was able to find μ_0 for a given value of n_1 or vice versa, by some cut-and-try method. For the integrals I_{n_1k} and $I_{n_1n_1}$ he simplifies his calculation so that it involves only the ratio

$$\frac{I_{n_1k}I_{n_1r}}{I_{n_1n_1}}=\frac{-2(2n_1+1)}{(k,n_1)(r,n_1)}(1-\mu_0{}^2)P_k(\mu_0)P_r(\mu_0)\frac{dn_1}{d\mu_0}, \tag{39}$$

where (k,n) denotes $(k^2+k-n_1{}^2-n_1)$. The coefficient $dn_1/d\mu_0$ is then evaluated either graphically or by means of a series expansion in terms of Legendre polynomials. Because of the manual operations involved, such a scheme of computation seems to be relatively unsatisfactory. Since our problem of finding the characteristic values and the characteristic functions is a typical one encountered in many boundary-value problems, we can make full use of the well-known Ritz method, which is again an application of a variational principle.

VI. APPLICATION OF RITZ METHOD TO DETERMINE n, AND THE CORRESPONDING CHARACTERISTIC FUNCTIONS

Without going into the details of the method, we will give only the essential steps pertinent to our problem, which is to find the characteristic value n of the

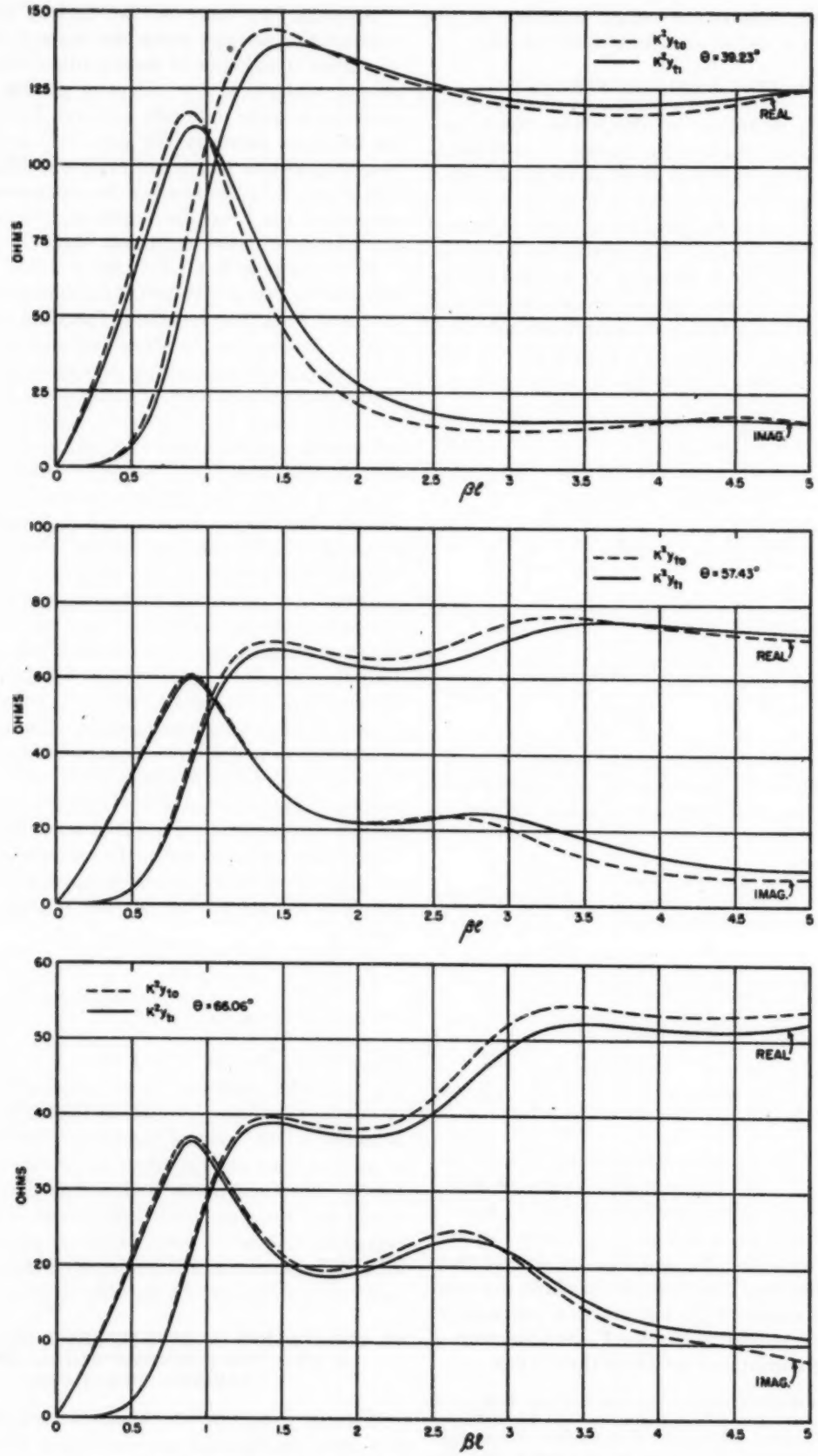


FIG. 4. Zeroth-order and first-order solutions for wide-angle cones.

Legendre equation

$$\frac{d}{d\mu}\left[(1-\mu^2)\frac{dL_n}{d\mu}\right]+n(n+1)L_n=0 \tag{40}$$

with the given boundary condition

$$L_n(\mu_0)=L_n(-\mu_0)=0; \quad \mu_0=\cos\theta_0, \tag{41}$$

where μ_0 is a prescribed value of μ corresponding to a given conical angle θ_0. In this formulation, we look only for those solutions L_n which are odd functions with respect to μ. Following the notation used by Madelung,[10] we let

$$L_n=\sum_{p=1}^{k} C_p f_p(\mu), \tag{42}$$

where $f_p(\mu)$ denotes a set of orthogonal functions properly normalized, with the arbitrary constants C_p to be determined later. We choose here

$$f_1=\frac{(3\cdot5\cdot7)^{\frac{1}{2}}}{4\mu_0^2(\mu_0)^{\frac{1}{2}}}\mu(\mu^2-\mu_0^2), \tag{43}$$

$$f_2=\frac{(3\cdot5\cdot7\cdot9\cdot11)^{\frac{1}{2}}}{8\mu_0^5(\mu_0)^{\frac{1}{2}}}\mu(\mu^2-\mu_0^2)(\mu^2-\tfrac{1}{3}\mu_0^2), \tag{44}$$

$$f_3=\frac{11\cdot13(5\cdot7\cdot9)^{\frac{1}{2}}}{32\mu_0^7(2\mu_0)^{\frac{1}{2}}}\mu(\mu^2-\mu_0^2)(\mu^2-\alpha\mu_0^2)(\mu^2-\beta\mu_0^2), \tag{45}$$

$$\alpha,\beta=\frac{1}{13}\left(5\pm\frac{2}{11}(220)^{\frac{1}{2}}\right).$$

...

It can be easily verified that

$$\int_{-\mu_0}^{\mu_0} f_p f_q d\mu=\delta_{pq}, \quad p,q=1,2,3\cdots \tag{46}$$

where δ_{pq} denotes the Kronecker delta. Higher order polynomials can be constructed, based upon the relations defined by Eq. (46). If we define the coefficient A_{pq} by the following integral

$$A_{pq}=\int_{-\mu_0}^{\mu_0}(1-\mu^2)\dot{f}_p\dot{f}_q d\mu, \tag{47}$$

where the dot denotes derivative of the function. Then the first few coefficients are given by

$$A_{11}=\frac{21}{2\mu_0^2}\left(1-\frac{11}{21}\mu_0^2\right), \tag{48}$$

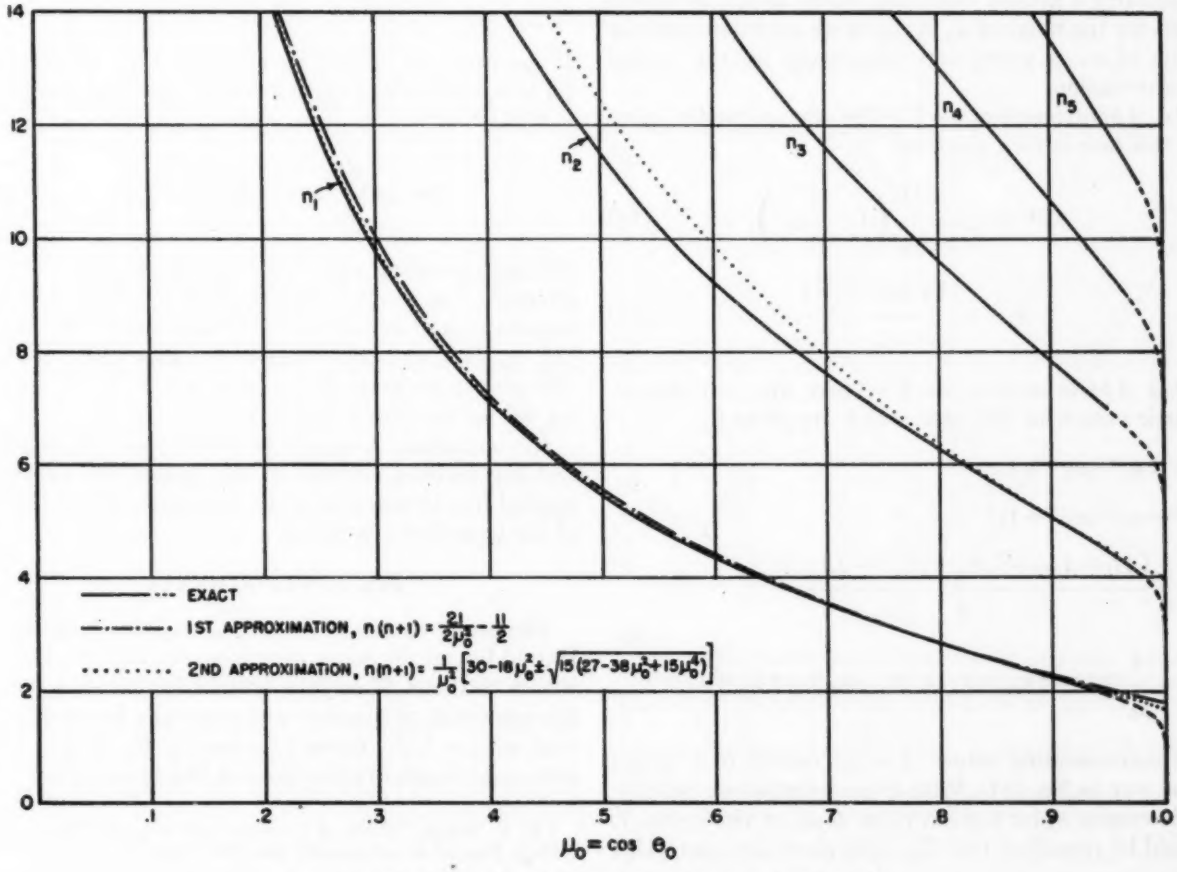


FIG. 5. Distribution of the characteristic values as a function of μ_0.

[10] E. Madelung, *Die Mathematischen Hilfsmittel des Physikers* (Dover Publications, New York, 1943), pp. 217–219.

TABLE I. First characteristic value of L_n as a function of μ_0.

μ_0	0.9376	0.774597	0.538469	0.405845	0.324253	0.269543	0.230458
$n_1^{(1)}$	2.08724	3	5.06446	7.14843	9.22712	11.30135	13.37263
$n_1^{(2)}$	2.01839	3	5	7.00032	9.00100	11.00100	13.00249
n_1^{exact}	2	3	5	7	9	11	13

$$A_{12}=A_{21}=\frac{3(11)^{\frac{1}{2}}}{2\mu_0^2}\left(1-\frac{5}{3}\mu_0^2\right), \tag{49}$$

$$A_{22}=\frac{99}{2\mu_0^2}\left(1-\frac{61}{99}\mu_0^2\right). \tag{50}$$

If we denote $n(n+1)$ by λ, then the coefficients C_p satisfy the following system of linear equations:

$$\sum_{p=1}^{k} C_p(A_{pq}-\lambda\delta_{pq})=0, \quad (q=1, 2\cdots k). \tag{51}$$

The characteristic values, k in number, are given approximately by the roots of the following secular equation,

$$y(\lambda)=|A_{pq}-\lambda\delta_{pq}|=0, \quad (p, q=1, 2\cdots k). \tag{52}$$

The method is simple in theory, but becomes impractical if we carry k greater than, say, three. As will be shown later, for the range of μ_0 in which we are interested, the value of n_1 is given very accurately by the second approximation.

1st Approximation, $k=1$.—The characteristic value for this case is then given by

$$\lambda_1^{(1)}=A_{11}=\frac{21}{2\mu_0^2}\left(1-\frac{11}{21}\mu_0^2\right) \tag{53}$$

or

$$n_1^{(1)}=\frac{(1+4\lambda_1^{(1)})^{\frac{1}{2}}-1}{2}. \tag{54}$$

2nd Approximation, $k=2$.—There are two characteristic values for this case, which are given by

$$\left.\begin{aligned}\lambda_1^{(2)}&=n_1^{(2)}(n_1^{(2)}+1)\\ \lambda_2^{(2)}&=n_2^{(2)}(n_2^{(2)}+1)\end{aligned}\right\}$$

$$=\frac{(A_{11}+A_{22})\mp((A_{11}-A_{22})^2+4A_{12}^2)^{\frac{1}{2}}}{2},$$

$$=\frac{1}{\mu_0^2}[(30-18\mu_0^2)\mp(15(27-38\mu_0^2+15\mu_0^4))^{\frac{1}{2}}]. \tag{55}$$

The corresponding values of n are related to λ in the same way as Eq. (54). With these formulas we are able to determine n_1 for a given value of μ_0, or vice versa. It should be remarked that Eq. (53) gives the *exact* value of n, namely 3, when μ_0 is equal to $\pm(\frac{3}{5})^{\frac{1}{2}}$, which is the non-vanishing root of $P_3(\mu_0)=0$. For this particular case the function $f_1(\mu)$ is the same as $P_3(\mu)$ except that they are normalized differently. The second approximation gives two *exact* characteristic values of n, namely 3 and 5, when μ_0 is a root of $P_3(\mu)$ or $P_5(\mu)$. The values of n, as a function of μ_0, as computed from Eqs. (53) and (55), are plotted in Fig. 5. The curve marked "exact" is constructed by joining the roots of the odd-order Legendre polynomials[11] and those of the even-order Legendre function of the second kind. In the neighborhood of $\mu_0=1$, n behaves like $k+1/\log(2/\theta_0)$, $k=1, 3\cdots$; therefore the approximate values of n deviate considerably from the exact values. Fortunately, for small-angle cones we have an exact expression for K^2Y_t and there is therefore no need to determine n in this region. It is obvious from Fig. 5, or from Table I, that the second approximation gives very accurate results for the value of n_1, provided μ_0 is not too near unity.

Equations (53)–(54) also provide us with a simple way to compute the coefficient $dn_1/d\mu_0$ which is connected with the integrals I_{n_1k} and $I_{n_1n_1}$ through Eq. (39). If we use Eq. (53), then the coefficient is simply given by

$$\frac{dn_1^{(1)}}{d\mu_0}=\frac{-21}{(2n_1^{(1)}+1)\mu_0^3}. \tag{56}$$

It can easily be verified by means of Eqs. (37–39) that for $n_1=3$ the value thus obtained is *exact* and is equal to $-5(5/3)^{\frac{1}{2}}$, or

$$(1-\mu_0^2)\frac{dn_1}{d\mu_0}=-2(5/3)^{\frac{1}{2}}, \quad n_1=3. \tag{57}$$

The corresponding value obtained by Smith (see Fig. 3, reference 3) is about 5 percent off. If we divide both the numerator and denominator of the second term of Eq. (33) by $I_{n_1n_1}$ and then make use of Eqs. (39), (53) and (56) we can evaluate K^2Y_{t1} for $n_1=2$, 5/2, 7/2, etc., as we did for the case $n_1=3$, 5, 7.

In conclusion, it should be noted that the material and the method covered in this paper can be easily applied to a biconical acoustic horn with proper change of the boundary conditions.

ACKNOWLEDGMENTS

The writer wishes to express his indebtedness to Dr. Harold Levine for many valuable consultations. He also wishes to thank Miss Julie Klimas for carrying out all the numerical calculations and especially for evaluating with such a high degree of accuracy the roots of the even-order Legendre functions of the second kind.

[11] E. R. Smith, "Zeros of Legendre polynomials," Iowa State College Journal of Sciences **12**, 263–274 (1938).

Erosion of Electrical Contacts on Make

L. H. Germer and F. E. Haworth
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received May 6, 1949)

When an electric current is set up by bringing two contacts together, they necessarily discharge a capacity. If the discharge takes place through an arc before the metallic circuit is established, erosion of the electrodes results. In a low voltage circuit the occurrence of an arc is dependent upon the condition of the electrode surfaces and upon the circuit inductance. For "inactive" surfaces, and a voltage of the order of 50, an arc does not occur if the inductance is greater than about 3 microhenries. Surfaces of various metals can be "activated" by vapors of certain unsaturated organic compounds, and in the active condition they give arcs even when the circuit inductance is greater than this limiting value by a factor of more than 10^3. The study of arcs between active metal surfaces is not yet complete, and most of this paper and all of the remainder of this abstract have to do with surfaces which are inactive.

When an arc occurs at the make of inactive metal surfaces, its energy, which in low voltage circuits of practical interest is drawn entirely from a charged condenser, is dissipated almost entirely upon the positive electrode, and melts out a crater intermediate in volume between the volume of metal which can be melted by the energy and that which can be vaporized. Some of the melted metal lands on the negative electrode and, with repeated operation, results in a mound of metal transferred from anode to cathode. This transfer is about 4×10^{-14} cc of metal per erg.

The arc voltage is of the order of 15. If the initial circuit potential is more than about 50 volts there may be more than one arc discharge, successive discharges being in opposite directions and resulting in the transfer of metal in opposite directions—always to the electrode which is negative.

INTRODUCTION

When a condenser is discharged by bringing together two electrodes, the contact surfaces suffer erosion. This can happen to contacts of a relay in the telephone system if a condenser has been connected across them for the purpose of minimizing the voltage rise when they are separated. Even when no condenser is used the natural capacity of electrical wires themselves is often, and perhaps usually, sufficient to do significant damage when an electrical circuit is completed. This erosion, caused by the discharge of a capacity as contacts are brought together, is thus of fundamental importance in the performance of relay contacts.

One readily discovers that the erosion is due primarily to metal being melted on the positive electrode by a discharge which occurs before the contacts touch. We are concerned here not only with empirical measurement of the erosion, but also with the nature of the discharge which produces it and how both are related to circuit parameters, ambient gas, and other variables. The conclusions which have been reached are based upon: 1. measurements of the metal transferred from one electrode to the other, 2. microscopic observation of electrode surfaces after a single discharge, 3. determination of the separation at which the discharge occurs, and 4. oscilloscopic records of the discharge.

1. MEASUREMENTS OF METAL TRANSFER

In our first experiments it was found that, when an electrical circuit is established a large number of times by bringing together metal electrodes, a pit is developed in the positive and a mound on the negative; metal is transferred from positive to negative. In studying this transfer it has been convenient to separate and bring together crossed wires of the contact metal under investigation about 60 times per second by means of a magnetic loudspeaker unit as described on pp. 918–919 of an earlier paper.[1] Each closure discharges a condenser which is recharged after the contacts have been separated, the charging and subsequent insulation of the condenser from the voltage source being accomplished by an auxiliary non-chattering mercury relay driven through an adjustable phase-shifter by the oscillator which actuates the experimental contacts.[a] The volume

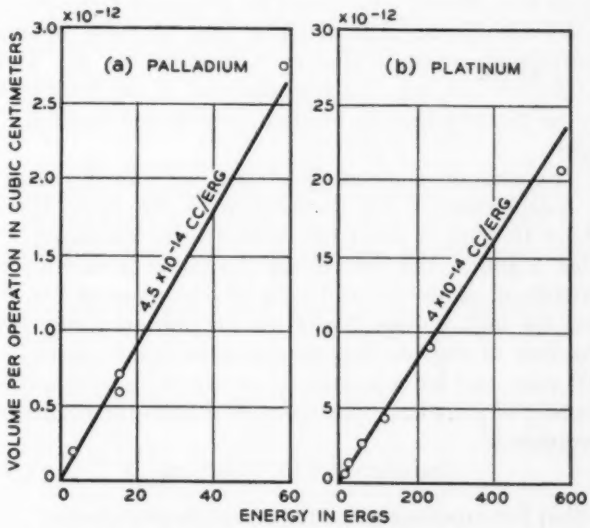


FIG. 1. Volume of metal transferred per condenser discharge, plotted against condenser energy. Number of closures varied from 4×10^5 at smallest energy to 0.25×10^5 at largest.

[1] J. J. Lander and L. H. Germer, J. App. Phys. 19, 910–928 (1948).

[a] This relay serves two purposes: (1) it eliminates the interruption of current at the experimental contacts when they are separating, which would also produce erosion, and (2) it keeps the charged condenser insulated from the voltage source when the contacts are brought together, so that possible chatter of the contacts on closure will not multiply the number of effective closures in arbitrary fashion and at the same time add erosion due to current interruption.

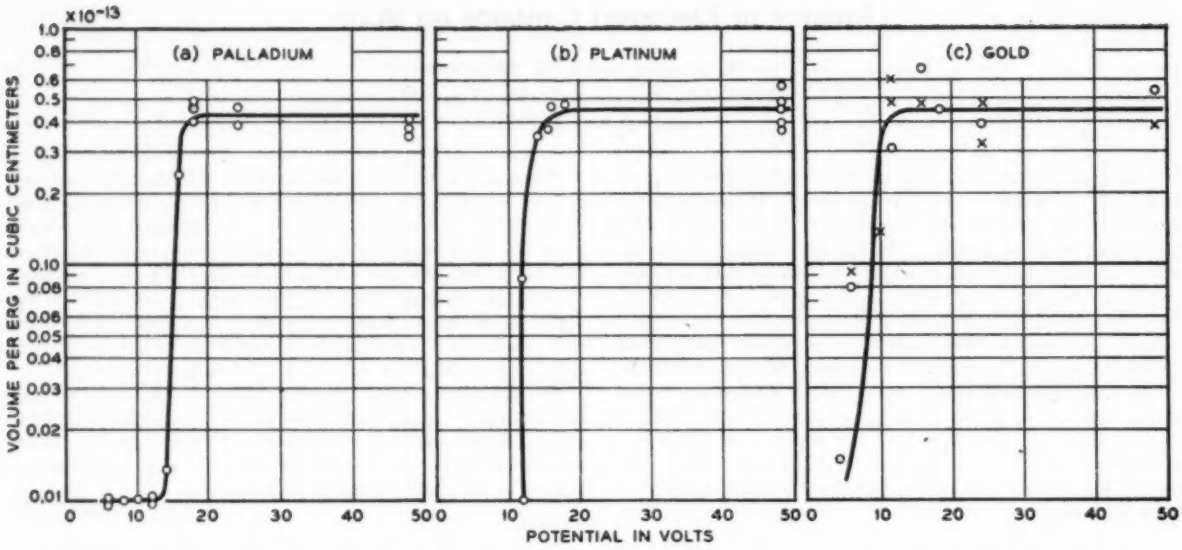


FIG. 2. Volume of metal transferred per erg of condenser energy plotted against condenser potential, in the low voltage range. (Transfer undetectably small is plotted on the 10^{-15}-cm³/erg line. Most experiments at 10^5 closures and about 11 ergs.)

of the mound which develops on the negative electrode after many closures is readily measured by the optical methods described on p. 919 of the earlier paper.[1]

In Figs. 1(a) and 1(b) are plotted as ordinates the volumes of metal transferred to the negative electrode per operation against the energy in the capacity which is being discharged, respectively, for palladium contacts in a 24-volt circuit and for platinum contacts in a 48-volt circuit. The straight lines of these figures correspond, respectively, to a transfer of 4.5 and 4.0×10^{-14} cm³/erg, but the precision of the tests was by no means adequate to justify the second significant figure.

The amount of metal transferred per unit of energy is independent of the potential over the range from about 15 volts to about 100 volts. This is illustrated by Figs. 2 and 3; the low voltage limitation is shown for palladium, platinum, and gold by the plots of Fig. 2 and the high voltage limitation for platinum only by the plot of Fig. 3. For the potential range of 15 to 100 volts, and for capacities which are not too high, the transfer of palladium, platinum, or gold to the negative electrode is

$$\text{transfer} = 4\times10^{-14}\ \text{cm}^3/\text{erg} \qquad (1)$$

within the experimental error of our measurements.

There is immediate interest in comparing the amount of metal transferred per erg with the quantity of metal which can be vaporized by an erg. In Table I are written down the calculated fractions of a cubic centimeter of each of the metals[2] palladium, platinum, and gold melted by one erg (column 2), raised to the boiling point (column 3), and vaporized (column 4). The volume of each of the metals actually transferred per unit of energy, 4×10^{-14} cm³/erg, is smaller than the volume which can be evaporated by unit energy (column 4, data for palladium not available) by the factor in the last column. One sees that the energy required to vaporize the metal which reaches the cathode is two or three percent of the total energy of the condenser.

TABLE I. Efficiency of condenser energy in producing transfer. Calculated cm³/erg (from the temperature of 300°K).

(1)	Melted (2)	To boiling point (3)	Vaporized (4)	Efficiency in producing transfer (5)
Pd	14×10^{-12}	7×10^{-12}	—	—
Pt	12	6	1.3×10^{-12}	0.03
Au	26	13	2.3	0.02

Transfer of metal is reduced by resistance or inductance in the circuit between the condenser and the contacts, where it limits the discharge current from the condenser. The two sets of points on the plot of Fig. 4 represent two sets of measurements of the transfer of platinum, for various resistances in a 48-volt 11-erg circuit. From measurements described in the next section one can deduce that with inductance rather than resistance a curve would be obtained essentially identical with that of Fig. 4, each 15 ohms on the scale of abscissas being replaced by 10^{-6} henry. The general shape of the curve of Fig. 4 is in agreement with our intuitive ideas, but its quantitative interpretation need not concern us here. In estimating the transfer of metal to be expected in any particular circumstance one must, one concludes from these experiments, take into account only that portion of the total circuit capacity which lies between the contacts and the nearest considerable resistance or inductance.

The operation of any actual relay differs from the

[2] Data from L. F. Vines, *The Platinum Metals and Their Alloys* (The International Nickel Company, Inc., New York, 1941), and from Int. Crit. Tab.

experiments described here in that a steady current is established upon closure of the contacts. Tests have been carried out to find the effect of this current upon the erosion. In these tests it was discovered that, in the absence of chatter, currents up to at least 1 amp. have no observable effect upon the erosion on make; the amount of metal transferred on make is small when the capacity is small even when the current is as large as 1 amp., and for currents up to 1 amp. it is determined by capacity and voltage only, being independent of current.

The proportionality between volume of metal transferred and condenser energy fails at low and high potentials, Figs. 2 and 3, at sufficiently high energies (of the order of 10^4 ergs), and when currents are established which are considerably larger than 1 amp. The reasons for the failures of the proportionality at low and at high potentials, and at large currents, will appear below. The failure at high energy will not be discussed here.

2. SINGLE DISCHARGE—MICROSCOPIC OBSERVATIONS

The results of the transfer experiments naturally lead one to examine the surfaces of contacts which have been brought together only once. The transfer measurements suggest also that the phenomena may perhaps be simplest when the potential across the condenser which is being discharged lies in the range between 15

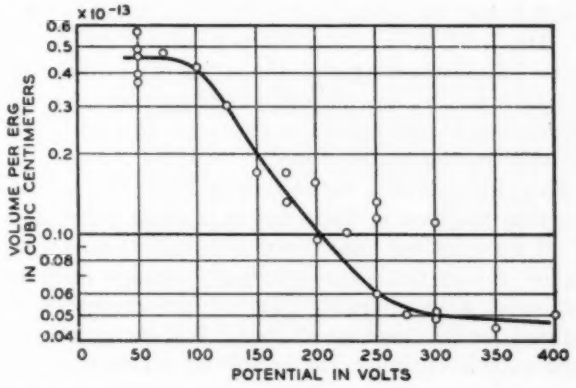


FIG. 3. Volume of platinum transferred per erg, in the high voltage range. (All experiments for 10^5 closures, and about 11 ergs.)

and 100 volts, and when the condenser energy is below 10^4 ergs.

When a condenser, charged to a potential between 15 and 100 volts, is discharged by bringing together a pair of electrodes, a pit is produced on the positive electrode with most of the metal from this pit forming a rim around it and with a small fraction of the metal spattered over on to the negative electrode. It is this spattered metal which accounts for the transfer from positive to negative on repeated operation. A photomicrograph of a pit formed on a positive platinum electrode by the discharge of an 0.001-mf condenser charged to 48 volts has been published.[3] The average diameter of this pit is about 5×10^{-4} cm and its depth is about one-quarter of its diameter. (Other such photographs appear in Fig. 6.)

Special precautions must be taken to insure a result as clean-cut as that just described. If contacts are brought together roughly, there is, in general, sufficient indentation of the surfaces to obliterate the pit completely. Even when the touching of the contacts is very gentle, one still may obtain a result which is not satisfactory. Multiple and overlapping pits are formed, in general, because the condenser is discharged several times after it has been repeatedly recharged from the voltage source; the recharging may occur during opening of the contacts on "chatter," which is difficult to eliminate, or, if the resistance in series with the battery and the velocity of approach are appropriate, recharge and redischarge may occur before the very first physical contact. Results are made ambiguous also by erosion produced by the interruption of current on chatter.

Moderately satisfactory experimental conditions have been obtained by charging the condenser attached to the contacts and then insulating it from the voltage source before they are brought together. Great precautions must be taken to have adequate condenser insulation and to bring the contacts together very gently. Although these precautions are not impossibly difficult, an easier experimental procedure has been devised. By this easier technique one contact is moved toward the other very slowly by means of a micrometer screw pushing the end of a cantilever bar, the movable contact being attached to the bar at a point which gives a reduction of the micrometer motion by a factor of 93 to 1. During the motion, the condenser connected across the contacts is attached to the source of potential through an appropriate resistance, which in many cases is 10^6 ohms, and a thyratron tube is so connected that any decrease in condenser potential greater than a pre-

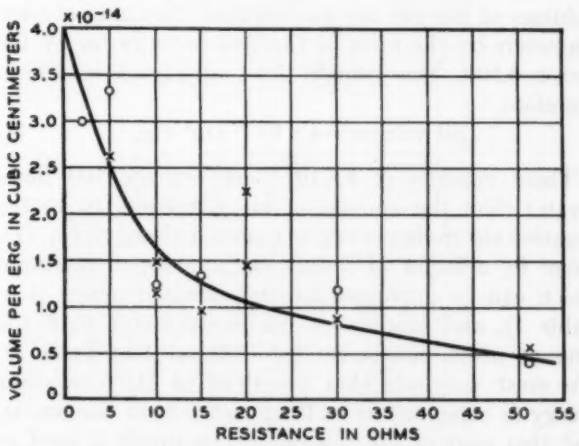


FIG. 4. Volume of platinum transferred per erg, with resistance in the condenser circuit. (All experiments at 10^5 closures, and 11 ergs in a 48-volt circuit.)

[3] L. H. Germer and F. E. Haworth, Phys. Rev. 73, 1121 (1948).

determined and adjustable amount disconnects the voltage source in a time of the order of 10 microsec.

The two sets of data of Fig. 5 represent the cubes of the average diameters of pits produced on positive platinum contacts plotted against condenser capacity in fractions of a microfarad, respectively, for 48 volts and for 24 volts. The two straight lines drawn through the experimental points correspond to 3.8×10^{-11} cm³/erg. If each of the pits has a depth equal to one-fourth of its diameter, which has been determined to be approximately true for the larger pits, and if we

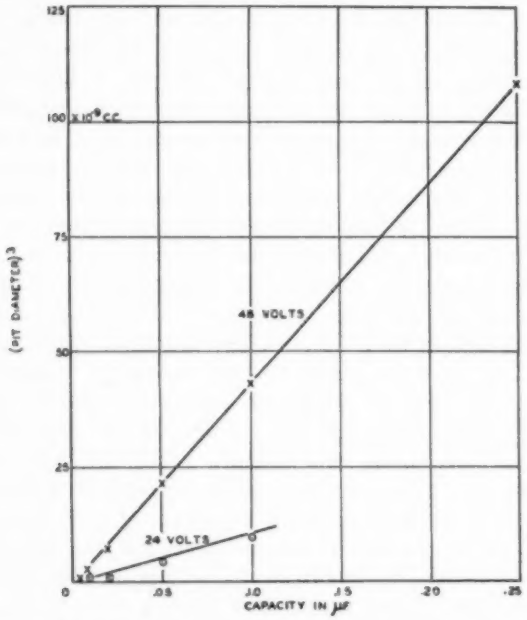


FIG. 5. Relation between pit diameter d for platinum and condenser capacity, for 48 and 24 volts. The straight lines correspond to $d^3=3.8\times10^{-11}$ cm³/erg, or approximately to volume$=4.0\times10^{-12}$ cm³/erg.

assume each pit to be a spherical segment, then the volumes of the pits are smaller than the cubes of their diameters by the ratio of $13\pi/384$ to unity, or by the factor 0.106. The straight lines of Fig. 5 represent, therefore,

$$\text{pit volume}=4\times10^{-12}\ \text{cm}^3/\text{erg}. \qquad (2)$$

These volumes of 4×10^{-12} cm³/erg are 100 times greater than the amount of metal transferred to the negative electrode per erg as reported above in Eq. (1), larger by a factor of 3 than the amount of platinum which can be vaporized by the energy (column 4 of Table I), and smaller by this same factor than the amount which can be melted (column 2 of Table I). We must conclude that one-third of the condenser energy is expended in melting metal upon the anode, and that part of the remaining two-thirds is used in vaporizing some of this metal. The consistency of the data certainly suggests strongly that not only some but *most* of the remaining two-thirds goes into heating metal above the melting point; most of the condenser energy is given to the anode and relatively little of this is lost by thermal conduction.

The fact that most of the energy is dissipated upon the anode suggests that the condenser is discharged before physical contact of the electrodes, by an electron current which bombards the anode. If the duration of the current were considerable the energy would be carried away by conduction in the anode metal and no appreciable heating would occur. The conclusion that very little of the energy is taken away by conduction places an upper limit on the duration of the discharge current. It is not easy to work out this limit precisely but a rough approximation is obtained without much difficulty.

For this calculation it is assumed for the sake of simplicity that the energy is dissipated at a single point on the anode and that the rate of dissipation is constant for the duration of the discharge. From a calculation on p. 923 of an earlier paper,[1] it is found that on these assumptions the temperature at the time t from the beginning of the discharge, at a distance r from the source of energy, is $T=(w/2\pi kr)(1-\text{erf}\ cr/t^{\frac{1}{2}})$, where w is the steady value of the power in watts, k is thermal conductivity, $c=\sigma^{\frac{1}{2}}h^{\frac{1}{2}}/2k^{\frac{1}{2}}$, σ is density, and h is thermal capacity in joules per °C per gram. For platinum c is about 1.0. The total heat loss from a hemisphere of radius r in the time t_1 is obtained by integrating $(2\pi r^2)k\partial T/\partial r$ with respect to t from $t=0$ to $t=t_1$. Of the total heat put into the hemisphere, the fraction which is lost by conduction is this integral divided by wt_1. Dr. J. M. Richardson has worked out for us that, for $r=2.5\times10^{-4}$ cm which is the radius of a pit produced on platinum by an 11 erg discharge and for $c=1$ appropriate to platinum, at a time as short as 2.5×10^{-7} sec. this fraction is about 0.5.

This calculation is very crude indeed. The assumption that the heat is dissipated at a single point is not at all right as we know from the fact that pits are shallow rather than hemispherical; perhaps more importantly, a considerable part of the energy is accounted for by the heat of fusion but the calculation neglects this entirely; the energy is furthermore not dissipated upon the anode at a constant rate, as we shall see later. In spite of the shortcomings of the assumptions involved in this calculation, its result certainly proves that the energy is dissipated upon the anode in an extremely short time, of the order of 10^{-7} sec. or less.

As soon as the conclusion has been reached that the condenser energy must be dissipated in a time of the order of 10^{-7} sec. or less, it is natural to inquire what sort of a *lower* limit is placed upon this time by the circuit parameters. Assuming that the circuit behaves as if it consisted only of a fixed resistance R in series with an inductance L and the condenser C, the discharge time is $\pi/[1/LC-R^2/4L^2]^{\frac{1}{2}}$. If R is negligible and $C=0.001$ microfarad, as in the actual case of the 11-erg discharge at 48 volts, a discharge time of 1×10^{-7}

sec. corresponds to $L=1\times10^{-6}$ henry. The circuit inductance in these experiments was actually less than this value by a factor of about 20. As a working hypothesis it seems reasonable to guess that the discharge time is really determined solely by the fixed circuit parameters, and it will be proved later that this hypothesis is true.

In the next section are reported direct measurements of the electrode separation at which the condenser is discharged. This distance is found to vary greatly in different experiments. For a potential of 50 volts across the condenser the discharge distances lie in the range between 1 and 15×10^{-5} cm. Now in most of the experiments reported here the electrodes are approaching each other at a velocity of the order of 1 cm/sec., at which velocity these critical distances are covered in times lying between 1 and 15×10^{-5} sec. Even the shortest of these times is greater by a factor of 10^2 than the 10^{-7} sec. upper limit for the duration of the discharge as calculated above. One must conclude that the discharge is a catastrophic phenomenon which takes place in a very short time during which the electrode separation can be considered constant.

Experiments upon other electrode metals, in addition to platinum, are naturally of interest. In such experiments it has been found that the volume of the pit produced in a gold positive contact is about the same as the volume of the pit in a platinum contact for the same condenser voltage and capacity. The pit in the gold contact is, however, different in shape; it is deeper and narrower, being more nearly hemispherical than the platinum pit. From the data of Table I one might expect that, for the same energy, a gold pit would have a volume about twice that of a platinum pit. Deviation from this expectation might occur because of differences in melting points and thermal conductivities of the two metals resulting in more loss of heat by conduction in one metal than in the other, but in any case the predicted factor of 2 is scarcely beyond the experimental error of measurements of pit volumes. The difference between pit shapes for gold and platinum has been confirmed in a large number of tests. The most simple explanation of this difference is that in the case of a gold anode the energy of the discharge is concentrated more nearly at a point than in the case of platinum; movement of the point of energy dissipation over the surface during the duration of the discharge would result in a shallow pit, and continued concentration at a single place in a pit which is less shallow.

Kodachrome photo-micrographs have been obtained of a gold anode and a platinum cathode after an 11-erg discharge at 48 volts, and of a platinum anode and a gold cathode after a similar discharge. Black and white reproductions of these photographs appear as Fig. 6. In each case the photograph of the anode shows a pit of the color of the anode metal without any trace of the cathode metal being apparent, and the photograph of the cathode shows a considerable roughened area a part of which has the color of the anode metal. Each anode pit is surrounded by a rim which obviously con-

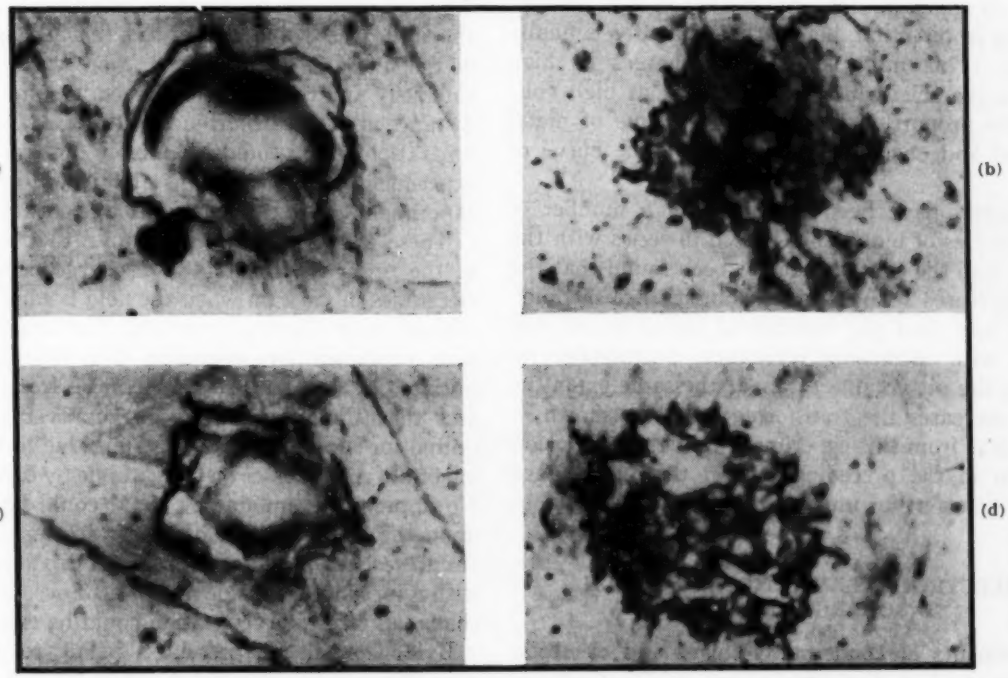


FIG. 6. Photo-micrographs of gold and platinum electrodes after a discharge of 11 ergs at 48 volts. Magnification 3700×. (a) Platinum anode which had been opposed to a gold cathode. (b) Gold cathode which had been opposed to a platinum anode. (c) Gold anode which had been opposed to a platinum cathode. (d) Platinum cathode which had been opposed to a gold anode.

TABLE II. Discharge voltages between polished contacts in vacuum.

Discharge	Silver contacts at 10×10^{-4} cm Volts	Calculated electric field (10^6 volts/cm)	Tungsten contacts at 5×10^{-4} cm Volts	Calculated electric field (10^6 volts/cm)
First	1360	1.36	740	1.48
Second	1420	1.42	820	1.64
Third	1700	1.70	1100	2.20
Fourth	1800	1.80	1780	3.56
Fifth	1820	1.82	2140	4.28
Sixth	1940	1.94	2380	4.76

tains most of the metal from the pit, of which only a small part is spattered over to the cathode. The difference between the shapes of gold and platinum pits is obvious on the reproductions of Fig. 6, the gold pit being deeper but less broad than the platinum pit, but naturally the two metals cannot be identified here by their characteristic colors.

For potentials below about 15 volts the pit formed on the anode is much smaller or non-existent. The sharp distinction between anode and cathode which is apparent in the photographs of Fig. 6 is, however, characteristic of low voltages only. After a discharge at 200 or 300 volts one cannot tell from the appearance of the electrodes which was anode and which was cathode; there is no longer a clearly marked pit on the anode. It will be proved later that the damage which is caused by the discharge is in most cases no longer markedly different for the two electrodes because at high voltages there is, in general, not one discharge but several, alternating in direction, before the condenser is finally discharged. This multiplicity of discharges at high voltages is the chief cause for the failure at high voltages of the proportionality between volume of metal transferred and energy, as exhibited by the curve of Fig. 3.

Pit volumes have been measured for a number of different values of inductance placed in series with the condenser and the contacts. It has been found that the volume decreases with increasing inductance in much the same way that the amount of metal transferred decreases with increasing resistance, 3 microhenries decreasing the pit volume to about the same extent as 50 ohms decreases the transferred metal as shown in Fig. 4. It was from this experiment that the conclusion was drawn in the preceding section that inductance decreases transfer in much the same way as does resistance.

3. ELECTRODE SEPARATION AT WHICH DISCHARGE OCCURS

Measurements of condenser voltages and electrode separations at which discharges occur are of importance in discovering the nature of the phenomenon. In experiments to be described here the polished ends of two wires are opposed to each other and the separation is adjustable by means of the micrometer screw and cantilever bar mentioned in the preceding section; the entire apparatus is mounted within a metal enclosure which can be exhausted and the micrometer adjustment is made by turning rods which extend through seals in the walls. The contacts can be observed through a glass window.

In a simple experiment, which is not at all new or original but is of interest because of its bearing on later work, two silver contacts were attached to a voltage source through a resistance of 10^6 ohms placed as close as possible to one of them. The separation was adjusted in a vacuum of 10^{-3} mm of mercury to 10×10^{-4} cm, and the potential between the contacts was then raised gradually while the gap between them was observed through a microscope at an over-all magnification of 125. Under these test conditions the mean free path of molecular ions was greater than the gap length by a factor of nearly 10^4; nevertheless, a discharge in the form of a small bright spark occurred when the potential reached 1360 volts. There was no accompanying observable change in the contact surfaces or in their separation. No further discharge took place until the potential had been increased to 1420 volts. After this, successive discharges took place at the progressively higher voltages given in Table II, corresponding to calculated electric fields rising from 1.4×10^6 volts/cm for the first discharge to 1.9×10^6 volts/cm for the sixth. In another experiment with tungsten electrodes at a separation of 5×10^{-4} cm the calculated field of successive discharges rose from 1.5×10^6 volts/cm for the first to 4.8×10^6 volts/cm for the sixth (see Table II). In this second experiment, as in the first, the contact separation was watched continuously and there was no detectable alteration in the distance upon the occurrence of the spark. The smallest separation which could be reliably observed under our experimental conditions was about 3×10^{-4} cm but no experiments were carried out at this separation.

These simple experiments prove that at potentials of the order of 10^3 volts a discharge occurs between electrodes in vacuum when the field is sufficiently high, without any physical contact. The observations are best explained by assuming that the discharge is initiated by field emission of electrons from the cathode and that the progressively higher average fields required for successive discharges are due to the smoothing out of roughnesses on the cathode by each discharge. No appreciable amount of metal could be thrown up upon the anode by the discharge because of the high circuit resistance and the absence of capacity in parallel with the contacts inside this resistance, and thus no change of electrode separation could be expected.

Evidence will be presented below to prove that, even at potentials very much below 10^3 volts, when two electrodes approach each other a discharge occurs before physical contact is established, although at lower voltages the separation has not been observed in

the microscope. It is to be presumed that the mechanism of the discharge is the same, in at least some of its characteristics, at low and at high potentials. At potential differences of the order of 10^3 volts results which seem significant for our present purposes are obtained only in vacuum. At 48 volts, discharge, as described below, occurs at a separation of the order of 0.9×10^{-4} cm; this distance is comparable with the mean free path of an electron in air at atmospheric pressure and the presence or absence of air seems to make no difference in the distance, as one might guess would be the fact. A great deal of experimentation has been carried out in vacuum for adequate reasons, but most of this will not be reported below. In most of the work to be described later in this paper the contacts were in air.

Most of our measurements of distances at which discharge occurs were made at 48 volts, and all of the following results were obtained at this voltage. This was chosen because it is in about the middle of the range over which the transfer of metal is proportional to the condenser energy, Figs. 2 and 3, and because it is the voltage used in telephone machine switching systems. It will be clear in a later section that 48 volts was a fortunate choice for the quantitative experiments we wished to make.

On the micrometer screw used for adjusting contact separation each division of the drum corresponds to 10^{-4} inch; the motion of the end of the screw is reduced by a factor of 93 by the cantilever bar, so that each division on the drum represents a change in separation of 273A. This is amply sensitive for measurements of the distance at which a 48-volt discharge occurs.

Great care is necessary in establishing the zero point at which the electrodes are just in contact. This was accomplished in the experiments reported below by closing and opening the contacts very gently in a circuit containing a potential of 15 millivolts, a microammeter, and a resistance of 10^3 ohms. With repeated operation in this low current and low voltage circuit a reliable zero point could finally be established, and after it was established the contacts were separated and then attached to a condenser charged to 48 volts. One side of the condenser was grounded and the other side was connected to the voltage source through a resistance of 10^6 ohms and through appropriate vacuum tubes, and connected also directly to the grid of a thyratron tube. The thyratron, together with the other tubes, performed the function of disconnecting the charging potential of 48 volts within a few microseconds after the discharge of the condenser had proceeded far enough to decrease its potential by as much as two or three volts, thus preventing the condenser from being recharged immediately. With this experimental arrangement the condenser was always fully charged when the contacts were being brought together, but at the same time whenever the discharge occurred the quantity of the charge was limited to that on the condenser plates without any appreciable supplement from the battery. After each measurement of discharge distance it was necessary to re-establish a zero point by repeatedly

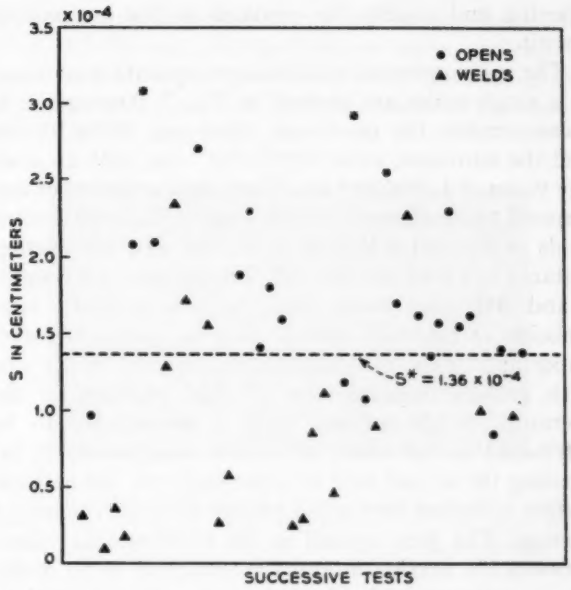


FIG. 7. Successive measurements of the electrode separation at which a condenser charged to 48 volts was discharged by platinum electrodes. Condenser capacity $13{,}000\times10^{-12}$f. Triangles indicate discharges after which the electrodes were in electrical contact, and circles those after which there was no contact.

TABLE III. Measurements of discharge distances at 48 volts.

Capacity ($\mu\mu$f) (1)	Number of discharges Resulting in Welds (2)	Opens (3)	Total (4)	Discharge distance (10^{-4} cm) Max. (5)	Aver.* (6)	Min. (7)	s^* (8)	Electric field (10^6 v/cm) Max. (9)	Aver. (10)	Min. (11)	Number of Opens below s^* (12)	Welds above s^* (13)
250	21	24	45	1.48	0.596	0.11	0.45	4.4	0.81	0.32	5	4
1000	20	23	43	1.54	0.605	0.030	0.47	16.0	0.79	0.31	6	4
3600	36	22	58	2.04	0.893	0.069	0.92	7.0	0.54	0.24	4	9
13,000	19	24	43	3.09	1.385	0.095	1.36	5.1	0.35	0.16	5	4
13,000	28	12	40	2.18	0.660	0.28	0.85	17.0	0.73	0.22	3	3
50,000	27	18	45	2.85	1.307	0.082	1.40	5.9	0.37	0.17	2	4
Total	151	123	274									
Average			45.2		0.91				0.60			

* One notes that there is a tendency for the average discharge distance to increase with increasing capacity. This is not fortuitous, but its occurrence is not significant. Some small control over surface smoothness was exercised in these experiments in order to have for each capacity an approximately equal number of opens and welds.

opening and closing the contacts in the low voltage circuit.

The measurements of discharge separation obtained in a single series are plotted as Fig. 7. Among the 43 measurements the maximum value was 3.09×10^{-4} cm and the minimum value 0.095×10^{-4} cm, with an average value of 1.38×10^{-4} cm. The extreme distances correspond to maximum and minimum calculated electric fields of 5.1 and 0.16×10^{6} volts/cm, and the average distance to a field of 0.35×10^{6}. The extremes are roughly 6 and 200 times lower than the field at which field emission of electrons from a cathode surface becomes important. This considerable discrepancy is in line with various measurements of field emission in the literature of the subject, and is presumably to be attributed to the effect of surface roughnesses in increasing the actual field at some spots on the cathode surface to values very much greater than the calculated average. The great spread of the experimental values between the maximum and the minimum is no doubt due to uncontrollable alterations in surface roughness and work function, caused by the discharges themselves and by the procedure used in establishing a reproducible zero point.

Some of the points of Fig. 7 are indicated by triangles and others by circles. These correspond, respectively, to experiments after which the electrodes were found to be in electrical contact, by test in the low voltage circuit, and to experiments after which they were not

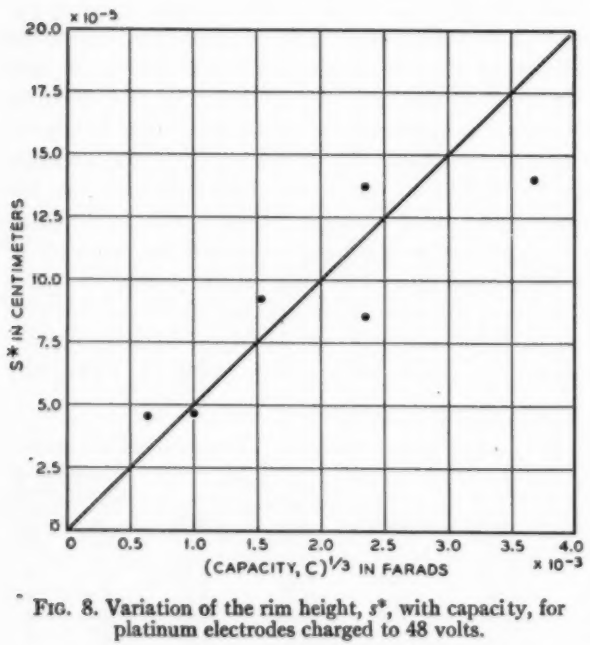


FIG. 8. Variation of the rim height, s^*, with capacity, for platinum electrodes charged to 48 volts.

in contact. One notes that the "opens" (circles) occur in general at greater distances than the "welds" (triangles). In fact the horizontal line drawn at 1.36×10^{-4} cm on the figure pretty well separates the "opens" from the "welds;" there are above this line 19 circles and 4 triangles, and below it 15 triangles and 5 circles. This distance of $s^*=1.36\times10^{-4}$ cm must be considered the average maximum height of the mound of metal which is thrown up about the pit which is formed on the anode, as it is certainly this mound of metal which makes contact between the electrodes whenever such contact occurs as a result of the discharge. The height of the mound must be expected to increase with the size of the pit, that is, with the capacity of the condenser. A set of measurements like those of Fig. 7 has been made at each of a number of different capacities and the value of the distance s^*, above which discharges end, in general, in "opens" and below which they end in "welds," is, in fact, discovered to increase with capacity. The data are collected in Table III. The values of s^* obtained in these tests increase from 0.45×10^{-4} cm (column 8) for a capcity of $C=250\times10^{-12}$ farad (column 1) to 1.40×10^{-4} cm for $50{,}000\times10^{-12}$ farad. In Fig. 8 are plotted values of s^* against values of $C^{\frac{1}{3}}$. The straight line of this figure, corresponding to

$$s^*/C^{\frac{1}{3}}=0.05 \text{ cm/farad}^{\frac{1}{3}}, \qquad (3)$$

fits the points within the very large experimental error. The reason for using values of $C^{\frac{1}{3}}$ as abscissas in Fig. 8 will appear presently.

It is interesting to compare the relation to be expected between s^* and C with that of the straight line of Fig. 8. In experiments of earlier sections it was discovered that most of the metal from a pit on the anode is deposited in a rim about the pit, and in making a simple estimate of the height of this rim no significant error will be introduced by the assumption that all the metal from the pit is deposited there. Then, on the fairly accurate approximation that the volume of the rim is equal to that of the pit V, the average cross section of the rim is related in a simple way to the energy of the condenser through Eq. (2) and through the plot of Fig. 5. Now it is natural to assume an average shape factor β for the cross section of the rim, such that the average width of the rim is greater than the average height by the factor $\beta(>1)$ and to consider also a factor $\alpha(>1)$ for the ratio of the maximum rim height s^* to the average height. Collecting these relations, one has

$V=\pi d s^{*2}\beta/\alpha^2$ cm^3	$\alpha>1, \beta>1$
$V=0.046C$ cm^3	which is Eq. (2) for 48 volts with C expressed in farads.
$d=0.76C^{\frac{1}{3}}$	which is the straight line of Fig. 5 for 48 volts.

These yield,

$$s^*/C^{\frac{1}{3}}=0.14\alpha/\beta^{\frac{1}{2}}.$$

Using the experimental value of $s^*/C^{\frac{1}{3}}=0.05$ from Fig. 8 one obtains

$$\beta^{\frac{1}{2}}/\alpha=2.8.$$

Thus $\beta>8$.

One cannot, of course, draw any very significant conclusions regarding the average shape of the rim about a

platinum pit from such simple and very crude experimentation and reasoning as that described here. One can conclude only that the experimental data lead to eminently reasonable values of the factors which were introduced to describe average rim contours. The fact that numerical results obtained in such widely different ways fit together so well is evidence that our view is correct regarding the way in which a single discharge can produce immediate contact closure. This concept of the metal thrown up from an anode pit bridging completely across to the other electrode will be used in the following section, and observations made there will, on the other hand, support this concept further.[4]

4. OSCILLOSCOPIC STUDIES

Oscilloscopic records of the potential across a pair of contacts, during and immediately after discharges between them, have yielded much information about the nature of the discharges. In particular, oscilloscopic studies prove that the discharge voltage is about 15, which was suggested in Section 1 by the observation that there is little or no transfer of metal from one electrode to the other below a voltage of this order of magnitude, and in studies of Section 2 by the smallness or absence of an anode pit below about 15 volts. Much other information also has been obtained from oscilloscopic records, making the oscilloscope our most important research tool.

In obtaining most of the oscilloscopic records the condenser connected across the contacts is charged and then insulated from the charging potential before the contacts are brought together, as described in the first paragraph of Section 1. The correct phase relation, adjustable by a phase-shifter through which an auxiliary relay is driven from the same oscillator which actuates the experimental contacts, is recognized by such an oscilloscopic record as that of Fig. 9. The battery potential is here 48 volts which is the potential across the condenser and contacts until the auxiliary relay disconnects the battery. The potential across the contacts then begins to decrease because of a leakage resistance of about 10^7 ohms in the oscilloscope probe, and closure finally takes place at about 38 volts. Because of the low sweep speed, no details of the closure phenomenon appear on this figure.

FIG. 9. Oscilloscopic record of the potential across a pair o contacts, to illustrate correct adjustment of the auxiliary relay so that the condenser across the contacts is insulated from the voltage source when closure occurs. Total sweep time about 10^{-2} sec.

A. Open Circuits

Many oscilloscopic records of the voltage across closing contacts have been taken with a moving film Fastax camera while the contacts are being opened and closed 60 times per second; the contacts are attached to a pair of deflecting plates of the oscilloscope through an adjustable attenuator-amplifier, and the oscilloscope sweep is not used. This method of operation has certain advantages because of the elimination of the sweep and the accompanying uncertainties, but the resolution is good to only about 2 microsec., which is not sufficient for some purposes.

The voltages across platinum contacts at four different closures are plotted against time in the Fastax camera photographs of Fig. 10. In each case a capacity of 1000×10^{-12} farad was connected across the contacts by short lead wires, the inductance of the circuit being measured to be about 0.05×10^{-6} henry. The condenser potential difference at discharge was 53 volts in the experiments of Figs. 10(a) and 10(b), and 145 volts in those of Figs. 10(c) and 10(d), these voltages being represented by the changes in the ordinates from the upper to the lower continuous lines. One observes no feature of interest in the closure of Fig. 10(a) but in the experiments of Figs. 10(b) and 10(c) a discharge between the contacts first discharged the condenser and then recharged it with the opposite polarity, the new potential persisting for a measurable time before closure; still more complex phenomena occurred in the test of Fig. 10(d). In Fig. 10(b) the reverse potential was -12 volt which lasted for 0.35 mm on the film, or a time of 12 microsec. as calculated from the measured film speed of 2950 cm/sec.; in Fig. 10(c) the reverse potential was -85 volt which lasted for 0.26 mm, or 15 microsec. as calculated from the film speed of 1700 cm/sec.; in Fig. 10(d) an intermediate potential of $+47$

[4] Early experiments by C. Kinsley ["Short spark-discharges," Phil. Mag. 9, 692–703 (1905)] resemble in some respects those reported in this section. Kinsley studied discharges of a 1-mf condenser between highly polished platinum-iridium surfaces at very small separations. He found that discharge took place whenever the quotient of applied voltage and distance reached a critical value. This was independent of voltage up to 130 volts, the highest applied, in a single series of tests but varied on different days from 1.2×10^6 to 8×10^6 volts/cm. After each discharge it was found that the surface had "cohered," but Kinsley states that "it is possible to prevent coherence by using a sufficiently small condenser."

Kinsley's measurements were more consistent than we have been able to obtain due probably to more carefully polished surfaces and to a more stable mechanical arrangement, and perhaps also to his policy of discarding individual distance determinations which seemed out of line. His maximum and minimum separations for a discharge at 48 volts were, respectively, 0.40 and 0.06×10^{-4} cm which are well within the extremes which we have observed, Table III, columns 5 and 7. Our plot of Fig. 8 would predict $s^*=5\times10^{-4}$ cm for $C=1$ mf, much larger than any distances observed by Kinsley or by us, which certainly accounts for the fact that Kinsley observed that the contacts *always* "cohered."

volts lasted for 37 microsec., and then a potential of -10 volt lasted for 23 microsec. before final closure. These data are collected in Table IV.

From the observations we can draw some simple deductions, which may be considered tentative until confirmed by later tests. The discharges between the contacts which ended in the open circuits of Figs. 10(b) and 10(c) were half-period oscillations of the circuit made up of the condenser, the gap between the contacts, and the short wires connecting the contacts and the condenser; the duration of each discharge should therefore be $\pi(LC)^{\frac{1}{2}}$ sec. where $C=1000\times10^{-12}$ f and $L=$approx. 0.05×10^{-6} h. This time is 0.02×10^{-6} sec. which is much too short to be recorded on the oscilloscope trace, and we are therefore necessarily without any direct record of the potential across the gap during the time of the discharge.

It is to be presumed that the closure of Fig. 10(a) was accomplished by a discharge of the same nature, but in this case the mound of metal thrown up upon the anode was sufficiently high to effect immediate short-circuit of the contacts. From Eq. (3) one calculates that for 53 volts (close enough to 48 for practical purposes) and a condenser of 1000×10^{-12} f s^* should be 5000A; thus one concludes tentatively that the closure of Fig. 10(a) was accomplished by a discharge which occurred at a separation of less than 5000A, and that the discharge which gave rise to the negative potential of Fig. 10(b) took place at a separation greater than 5000A. As a matter of fact the data of column 6 in Table IV can be translated into distance from the fact that the contacts were known to be moving at about 1 cm/sec. at closure. The 12 μsec. of the second line correspond then to a distance of 1200A, and if this is added to $s^*=5000$A for this closure, one obtains 6200A as the contact separation at which the discharge of Fig. 10(b) occurred. This calculation is obviously exceedingly crude, but it is interesting in principle.

TABLE IV. Data from the oscilloscope traces of Fig. 10.

Fig. (1)	Initial voltage at closure V_0 (2)	Open circuit voltage (3)	Film speed (cm/sec.) (4)	Length of open circuit (mm) (5)	Time of open circuit (microsec.) (6)
10(a)	53	—	2950	—	—
10(b)	53	−12	2950	0.35	12
10(c)	145	−85	1700	0.26	15
10(d)	145	+47	2650	0.98	37
		−10	2650	0.61	23

Although the potential across the contacts during the discharge was not directly observed in the records of Fig. 10, this quantity v can be estimated indirectly. During the discharge, current flows from the condenser C through an inductance L and a resistance R and against the constant potential v. The final potential V_1 at the end of the discharge, column 3 of Table IV for the experiments of Figs. 10(b) and 10(c), is related to the potential before the beginning of the discharge V_0 of column 2, by the equation

$$V_1=v(1+f)-fV_0, \tag{4}$$

where

$$f=[1/(1-x)^{\frac{1}{2}}]\exp[-(\pi+\sin^{-1}x^{\frac{1}{2}})[x/(1-x)]^{\frac{1}{2}}],$$

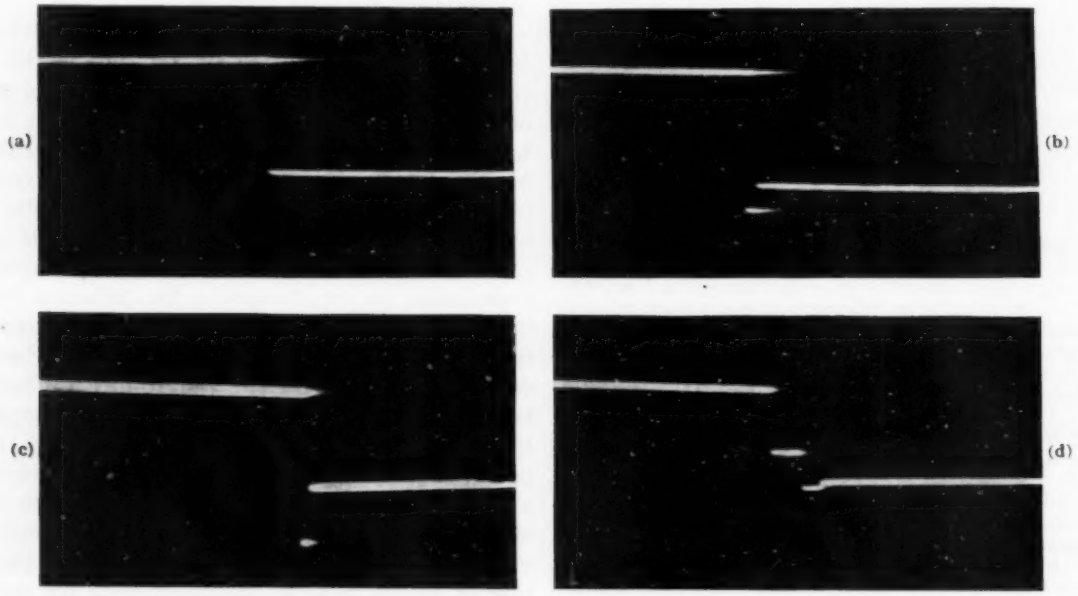


FIG. 10. Fastax camera records of potential across platinum contacts on closure. Length of reproduced trace 530 microsec. in (a) and (b), 920 microsec. in (c), and 590 microsec. in (d). (a) Closure at 53 volts without observable feature of interest. (b) Closure at 53 volts with intermediate open at −12 volts. (c) Closure at 145 volts with intermediate open at −85 volts. (d) Closure at 145 volts with intermediate opens at +47 volts and at −10 volts.

and $x=CR^2/4L$. The two independent pairs of values of V_0 and V_1 of Figs. 10(b) and 10(c) yield $v=16.8$ volts and $f=0.79$. This value of f corresponds to $x=0.0050$ or, for the known magnitudes $C=1000\times10^{-12}$ and $L=0.05\times10^{-6}$, to $R=1.0$ ohm. This is reasonable for the high frequency resistance of the circuit.

The experimental value of the potential v is the most interesting datum obtained from the traces of Figs. 10(b) and 10(c). To find out whether or not experiments carried out at other initial potentials are consistent, a series of photographs has been obtained at each of 78 different initial potentials with identical circuit conditions. Of 7251 photographs, 4125 exhibit intermediate opens like those of Figs. 10(b), 10(c), and 10(d), and the remainder show closures without features of interest. In Fig. 11 the open circuit potentials (the first if there are more than one) of all of these 4125 traces are plotted as ordinates against values of V_0, the initial potential. Many points of this plot fall along a straight line which fits Eq. (4) with

$$v=15 \text{ volts and } f=0.85;$$

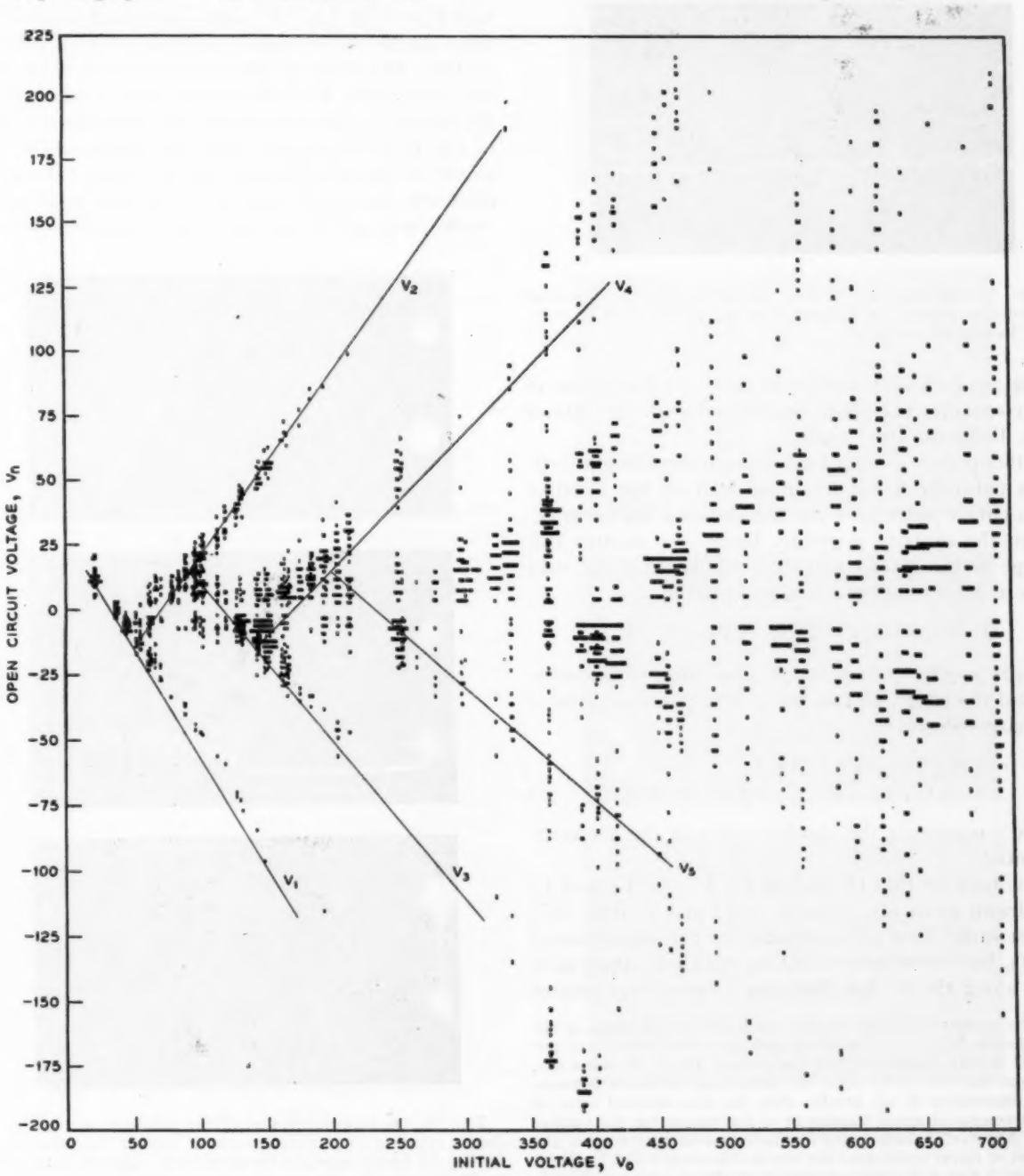


FIG. 11. Plot of open circuit potentials, V_n, as ordinates against initial potentials, V_0, as abscissas. The straight lines represent Eq. (5) for different values of n, with $v=15$ volts and $f=0.85$.

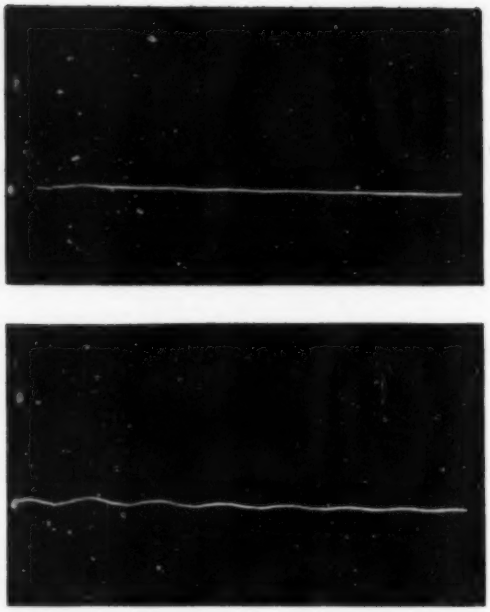


FIG. 12. Oscilloscopic traces of the potential across platinum contacts on closure. (a) Closure after an arc at 15 to 18 volts. (b) Closure without an arc.

these are then presumed to be more reliable values of the constants than those determined from the data of Figs. 10(b) and 10(c) alone.[b]

Other points of Fig. 11 fall along other straight lines. It is naturally to be presumed that, if the negative open circuit potential V_1 at the end of a discharge between the contacts is greatly below $-v$, another discharge in the reverse direction will be likely to take place at once to reach a final potential of

$$V_2=-v(1+f)-fV_1=-v(1+f)^2+f^2V_0.$$

If V_2 is positive and above $+v$, there may of course be a third discharge, and so forth. The potential after n discharges should be,

$$V_n=(-1)^n[f^nV_0-v(1+f)(1+f+f^2+\text{——}+f^{n-1})]. \quad (5)$$

(Here v represents the absolute value of the discharge potential.)

The lines on Fig. 11 marked V_1, V_2, V_3, V_4, and V_5 are drawn to fit Eq. (5) with $v=15$ and $f=0.85$. Still higher order lines are suggested by the experimental points, but the erratic variation, which is appreciable even along the V_1 line, becomes progressively greater so that it is unprofitable to draw in lines representing more than 5 successive discharges.

From $f=0.85$ one calculates $x=0.0030$ which corresponds to $R=0.77$ ohm for $L=0.05\times10^{-6}$ h., but since this latter value has considerable uncertainty the high frequency resistance is not known with any precision.

The oscilloscopic record of Fig. 10(d) is now readily interpretable. The first open circuit at +47 volts is the potential which was reached after two discharges in rapid succession, and the open at −10 volt the potential after a third discharge. If we set $V_0=145$, $V_2=+47$, and $V_3=-10$, Eq. (5) yields two simultaneous equations which can be solved to give $v=15.8$ volts and $f=0.83$. The slight disagreements between these values and those from Figs. 10(b) and 10(c) are examples of the erratic variation shown by the experimental points of Fig. 11. It will appear later that this variation is due chiefly to changes in the discharge voltage between one discharge and another and even during the course of a single discharge. The nature of this change is not under-

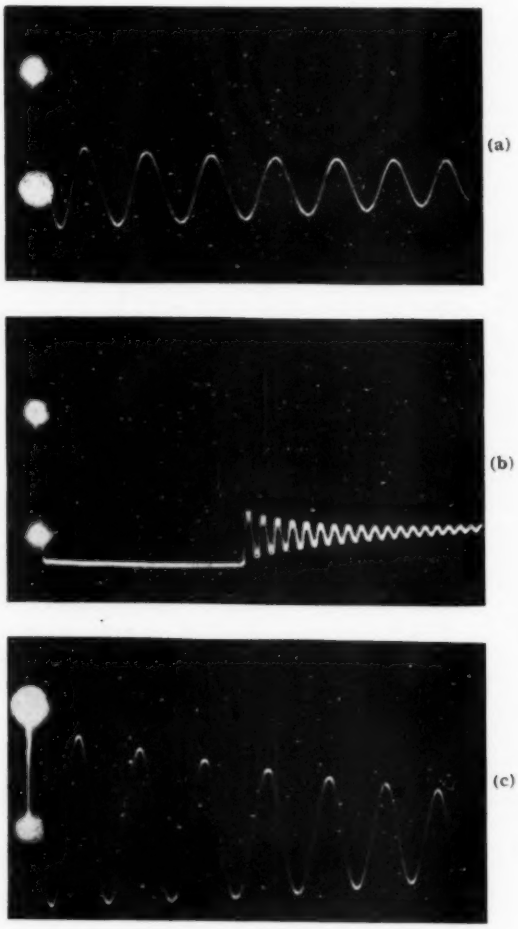


FIG. 13. Oscilloscopic traces of the potential across the condenser attached to platinum contacts. Circuit constants same as for Fig. 12. Sweep time also the same for (a) and (c), but increased to 50 microsec. for (b). (a) Closure of contacts after an arc. (b) Open circuit, after an arc—lasting for 23 microsec., then final closure. (c) Closure without an arc.

[b] The energy dissipated on the anode by the discharge is approximately $2C(V_0-v)v$, assuming zero circuit resistance, and not $CV_0^2/2$ as was concluded from the data of Fig. 5. It turns out, however, that for $v=15$ volts the percentage difference between these expressions is not greater than the experimental error of measurements of metal transfer or of pit volume for the voltage range $20<V_0<75$, accounting for the apparent constancy of the amount of metal transferred per erg in this range which is shown by Figs. 2 and 3, for the consistency of the two lines of Fig. 5, and for the consistency of the data upon discharge distance obtained at 48 volts.

stood. There is another cause for the scattering of certain points of Fig. 11 which *is* understood and will be discussed below in Section C.

Among all of the 4125 Fastax camera photographs which exhibit opens there are altogether 267 which resemble Fig. 10(d) in that each of them shows two opens. Of these there are two which exhibit three opens, the third of which is apparently also of the type under consideration, and in addition a very few other third opens which appear to be examples of the phenomenon described below in Section D.

With knowledge now available one readily estimates lower limits to the maximum current density in a discharge between contacts on make. The maximum current in a discharge is approximately $(V_0-v)(C/L)^{\frac{1}{2}}$, and the cross section of the discharge is certainly less than that of the pit formed on the anode. If the cross section were equal to this area the maximum current density in the 11-erg discharge which produced the typical platinum pit[3] of diameter 5×10^{-4} cm, for example, would have been 2.4×10^{7} amp./cm^2. It seems certain that the discharge must have been confined to an area smaller than the area of the anode pit, and that therefore the maximum current density in this discharge must have been greater than 2×10^{7} amp./cm^2, and probably of the order of 10^{8} amp./cm^2. Estimates of current density will turn out to be of value in working out a theory of the discharge.

In the first section it was reported that the erosion of contacts on make is due to the energy stored in a condenser, being independent of the steady value current which is set up unless this current is quite large. The reason for this lack of dependence upon steady value current is now clear. The erosion produced by the arc depends upon the total electrical charge passed by it, and this charge is drawn partly from the condenser and partly from the source of potential. The arc ends if, and when, the current becomes zero (or of course when the contacts touch). If the contacts are connected to the voltage source through leads of sufficiently low resistance and inductance the current will *never* become zero and the total charge passing through the arc may be increased many fold by the battery current, but unless the current fails to reach zero the battery current will have relatively little effect upon the total charge and upon the erosion. The relation between circuit parameters which just keeps the current from reaching zero has been worked out for some special cases and found to be in agreement with the observed conditions which lead to the failure of Eq. (1). (See also Appendix A.)

B. Arc Voltages and More Detailed Studies of Open Circuits

Indirect determination of discharge voltage from such data as those of Fig. 11 is not completely satisfactory because it presumes that the voltage remains constant

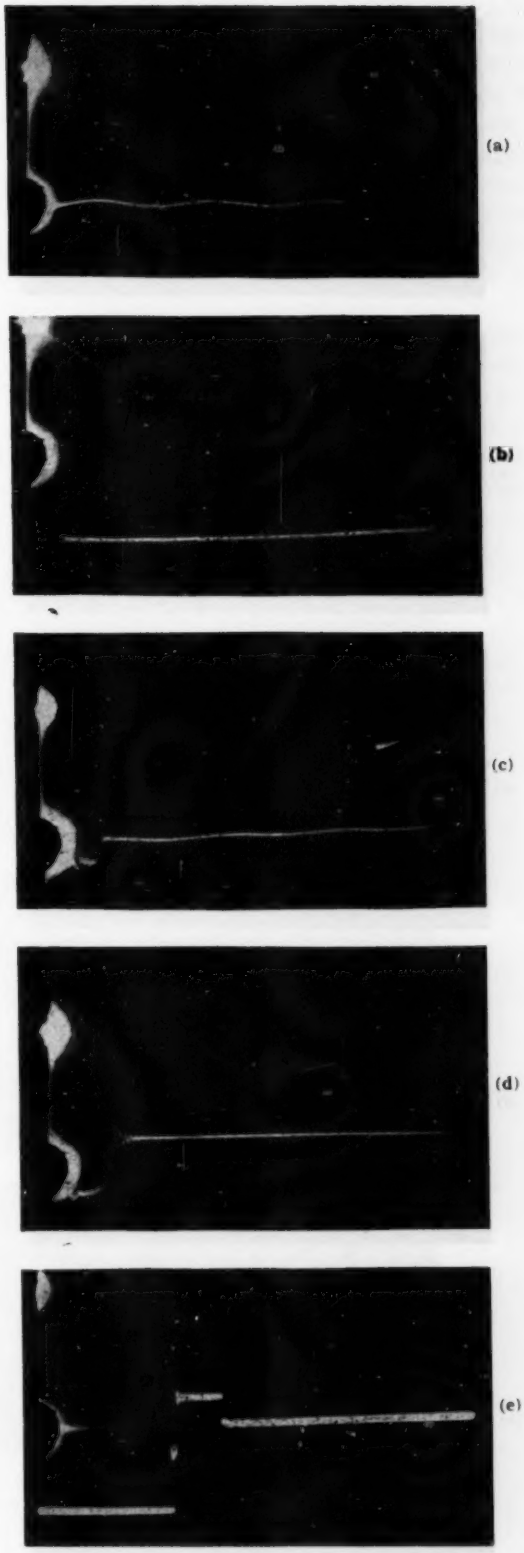


FIG. 14. Potential across platinum contacts on closure at 100 volts. Sweep time 4 microsec. for (a), (b), (c), and (d). (a) Closure after first arc. (b) Open after first arc. (c) Closure after second arc. (d) Open after second arc. (e) 40-microsec. sweep time—showing first open, second arc, second open, final closure.

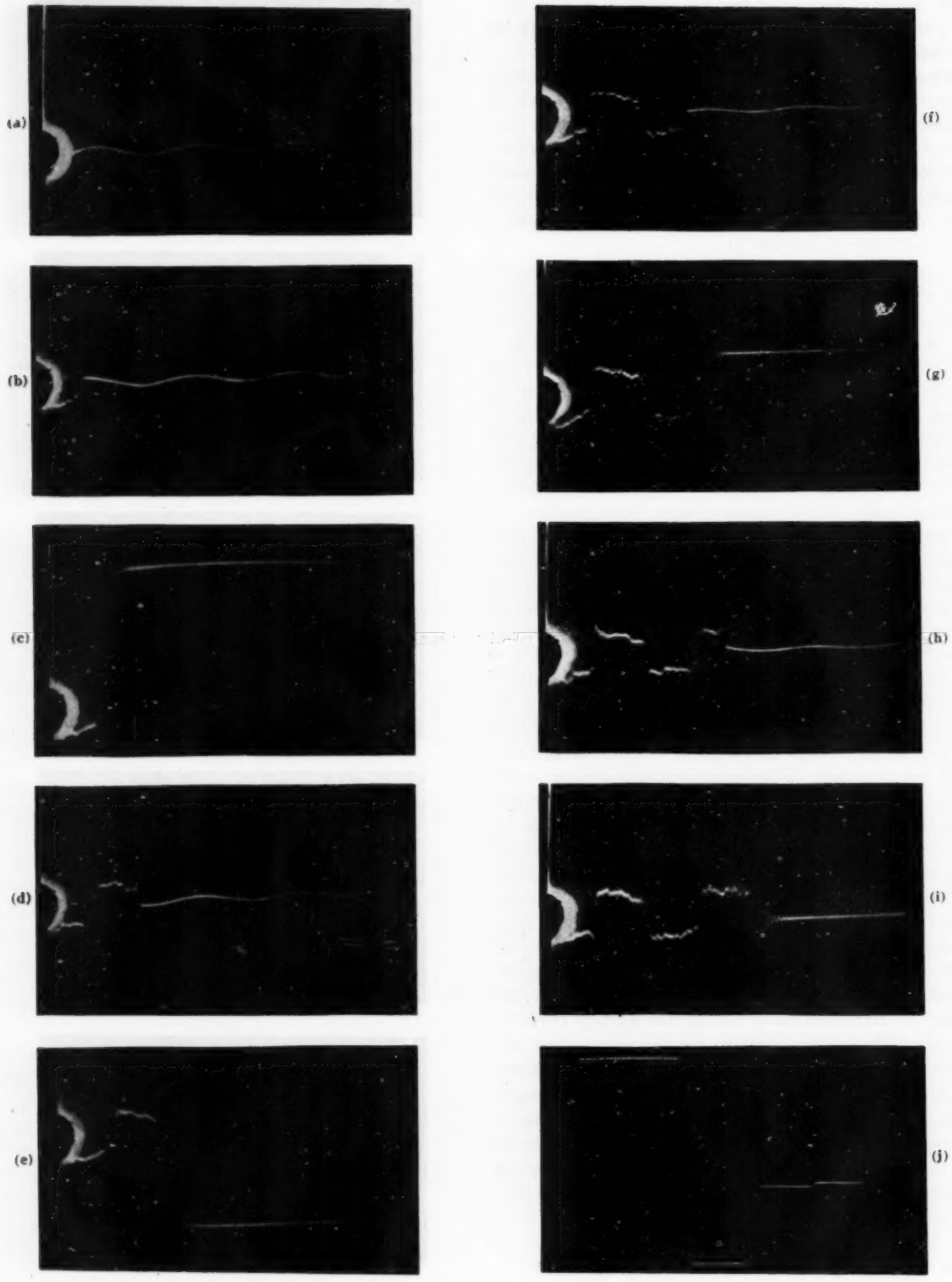


FIG. 15. Potential across platinum contacts on closure at 200 volts. Sweep time 4 microsec. for (A)–(I). (a) Closure after first arc. (b) Closure after second arc. (c) Open after second arc. (d) Closure after third arc. (e) Open after third arc. (f) Closure after fourth arc. (g) Open after fourth arc. (h) Closure after fifth arc. (i) Open after fifth arc. (j) Sweep time 40 microsec.—showing second arc, second open, third arc, third open, fourth arc, fifth arc, fifth open, final closure.

throughout each discharge, and because it fails to reveal important details of the phenomenon. Direct observation and measurement of the discharge potential is desirable, and this can be made if the period of the circuit is sufficiently increased.

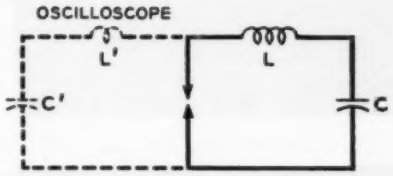


FIG. 16. The experimental circuit. The contacts are represented by arrows, the principal circuit which oscillates on closure by solid lines, and the oscilloscope with its leads by dashed lines.

An increase of the condenser from 1000 to 20,000 $\times 10^{-12}$ farad and the insertion of an inductance of 3×10^{-6} henry are appropriate, as the half-period is then 0.77×10^{-6} sec. which can be readily observed by using the oscilloscope sweep and photographing the screen in the normal manner. Photographs of closures under these conditions are reproduced in Fig. 12. When these were taken the contacts were being operated 60 times per second and the potential change at each closure tripped the oscilloscope sweep while the signal was supplied to the deflecting plates through a delay network giving a nominal delay of 0.5 microsec. The two bright spots at the left represent potentials of 0 and of 48 volts across the contacts. Each photograph was obtained by opening the camera shutter for 0.01 sec., thus insuring the recording of one trace only. The sweep time was 10 microsec. One sees in Fig. 12(a) that the potential across the contacts dropped from the initial value of 48 volts to something of the order of 15 or 18 at which it held in rather unsteady fashion for about half a microsecond and then fell further to zero. The discharge at a voltage

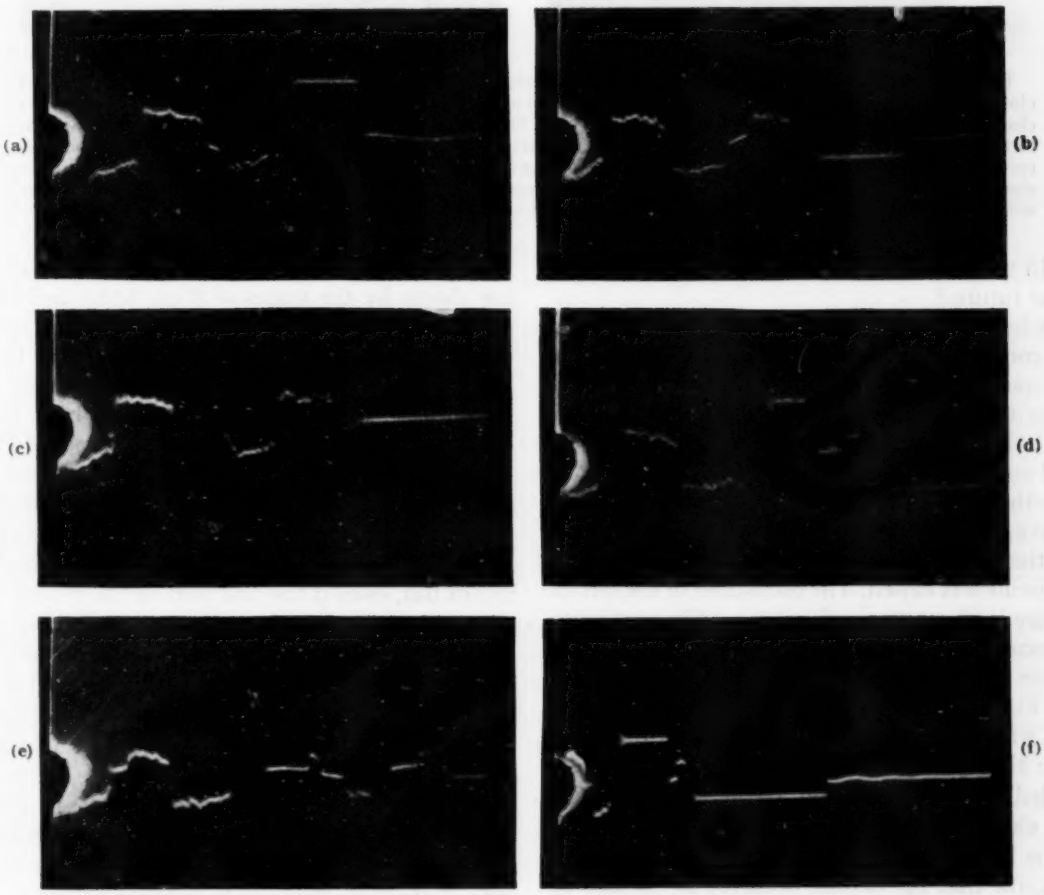


FIG. 17. Potential across platinum contacts showing "transient closures." Experimental circuit conditions like those of Fig. 15. Sweep time 4 microsec. for (a)–(e). (a) Transient closure after third arc, lasting only 20 percent of the time of the fourth arc, followed by the fourth arc, by a fourth open, and then by final closure. (b) A similar transient closure for the first part of the fifth arc, a fifth open, and final closure. (c) A third open, then a transient closure opening to a fourth arc, a fifth arc and a fifth open. (d) A fourth open, then a transient closure opening to a fifth arc, and at the end a fifth open. (e) Four transient closures, one in each of the third, fifth, sixth, and seventh arcs, ending in a final *seventh open*. The most elaborate pattern which has been recorded. (f) Sweep time 10 microsec. An extended fourth open, a transient closure opening to a fifth arc, an extended fifth open, and final closure.

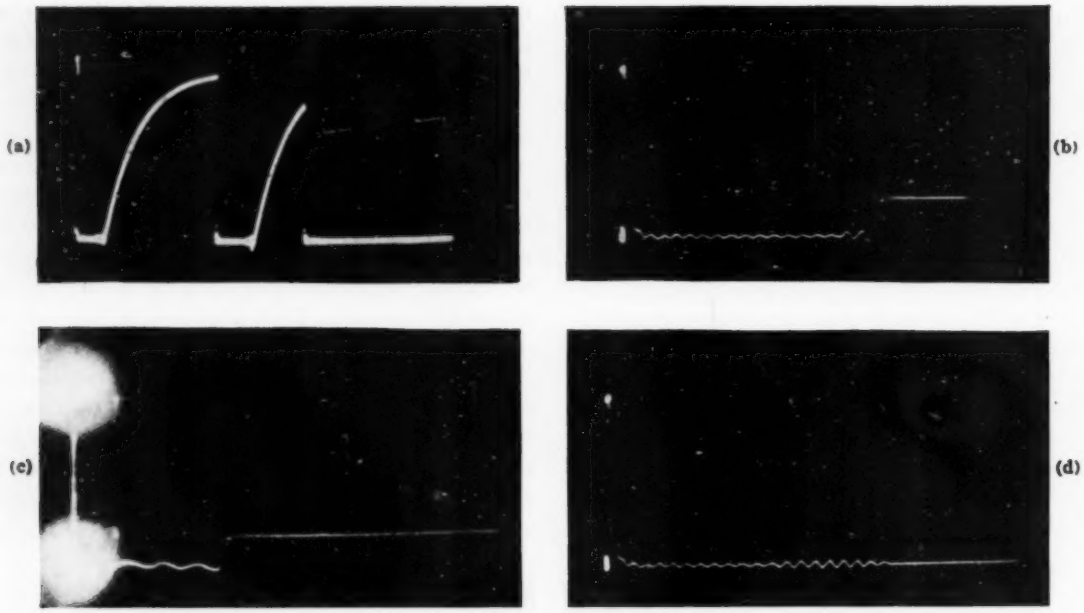


FIG. 18. Potential across platinum contacts with mechanical chatter. (a) Voltage source not disconnected at closure, sweep time 100 microsec. First mechanical open at about 7 microsec. (b) Voltage source disconnected at closure in the normal fashion, sweep time 10 microsec. Showing the mechanical open at 7 microsec. (c) Similar to (b) with 5-microsec. sweep time and delay in the start of the sweep, to permit measurement of the open circuit frequency, 12×10^6 cycles/sec. (d) After adjustment to a condition in which the electrodes start to open mechanically with marked increase of contact resistance, but then at 7 microsec. permanent closure occurs without complete separation. (This figure is a part of footnote c).

of 15 or 18 we have called an "arc" and shall so refer to it in the future.[5]

The discharge represented by the arc of Fig. 12(a) ended in contact closure rather than in open circuit, although open circuits occur frequently as we already know from the Fastax camera photographs. The evidence that the arc ended in a true closure of the contacts, and not in an open circuit with the voltage close to zero, is the ripple of small amplitude which one can just observe along the base line. This ripple has the period of the oscillating circuit and its existence proves that the circuit was closed. The oscillation of the circuit upon closure is of course much more readily seen when the oscilloscope probes are attached across the capacity rather than across the contacts. Such photographs appear as Fig. 13. The trace of Fig. 13(a) was initiated by an arc which ended in a closure, without open circuit, while that of Fig. 13(b) was initiated by an arc which ended in an open circuit lasting 23×10^{-6} sec. before the electrodes touched.

From the initial amplitude of the oscillation of Fig. 13(a) we know that the closure occurred at the end of an arc. Not every closure, however, comes about as the result of a discharge; there are, at times and under conditions which will be discussed later, closures which are not preceded by arcs. Examples of such closures are shown by the traces of Figs. 12(b) and 13(c). The great amplitude of the oscillation of Fig. 13(c) is conclusive evidence that it was not preceded by an arc. Repeated photographs of the voltage across the condenser with the circuit conditions of Figs. 13 exhibit oscillations upon initial closure of these two different amplitudes (Figs. 13(a) and 13(c)), without any examples of amplitudes of intermediate magnitudes. The trace of Fig. 12(b) represents also a closure *without an arc;* on this trace the absence of an arc is obvious on the record but, even if the first part of the record had been obscured as it might have been if the initiation of the sweep had been slightly delayed or if the period of the circuit had been slightly shorter, one would still know that the closure was not preceded by an arc because of the considerable amplitude of the base line ripple in comparison with the ripple of Fig. 12(a). The two amplitudes of ripple of the patterns of Figs. 12 correspond respectively to the amplitudes of circuit oscillation of Figs. 13(a) and 13(c), and one can tell from the amplitude of the base line ripple whether or not an arc has preceded closure in cases in which the natural period of the circuit is so short that the arc is not observed directly.

The logarithmic decrement of the oscillation of Fig. 13(c) is measured to be about 0.10 which is approximately equal to $\pi R(C/L)^{\frac{1}{2}}$. From this one obtains $R=0.40$ ohm which is reasonable for the high frequency

[5] R. Holm, *Electric Contacts* (Hugo Gerbers, Stockholm, 1946), p. 261–263) has reported that at extremely small electrode separations the potential across an electric arc in air is of the order of 15 volts and is independent of the current, in agreement with our observations upon the discharge voltage.

resistance of the oscillating circuit. It is appreciably lower than the resistance deduced from the Fastax camera patterns of Fig. 11, presumably because it was measured at a considerably lower frequency.

If the initial voltage is much above 50, an oscilloscopic trace of the potential across a pair of contacts can show, in addition to an arc like that of Fig. 12(a), a second arc at a negative potential, indicating that a second discharge in the reverse direction occurred immediately after the first. When the voltage is sufficiently high several discharges may take place in succession, each being recorded by its appropriate voltage of about 15. Experimentation at potentials high enough to give more than one discharge is, in some ways, more satisfactory than at 50 volts because arcs beyond the first can be recorded completely, whereas the record of the first is never entirely accurate as the signal is used to trigger the sweep which then must reprodcue the signal itself. This is accomplished by the use of a signal delay network as mentioned above, but if one does not attempt to record the first arc no such network is necessary and there is a consequent gain in fidelity.

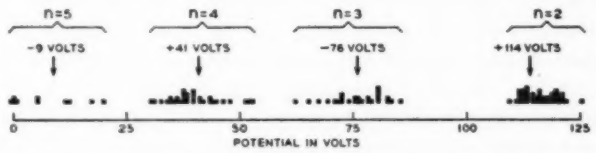


FIG. 19. Open circuit voltages of different orders for $V_0=200$ volts. Circuit inductance and capacity as in Figs. 15 and 17. The arrows are voltages calculated from Eq. (5) for $v=15, f=0.92$.

The traces of Fig. 14 record the potential across closing contacts for an initial voltage of 100 with a circuit capacity of $20{,}000\times10^{-12}$ f and an inductance of 1×10^{-6} h, and those of Fig. 15 for 200 volts and the same circuit constants. In obtaining these photographs the signal was used to initiate the sweep, but there was no delay network and the contacts were connected directly to the deflecting plates of the oscilloscope by fairly short wires.

In Figs. 14(a) and 14(b) one sees at the end of the first arc respectively a closure and an open circuit, the arc itself not being recorded at all because the sweep requires appreciable time to get underway and there was no delay network in the path of the signal. Figures 14(c) and 14(d) show a second arc followed, respectively, by a closure and by an open circuit. The final history of the open of Fig. 14(b) is a matter of conjecture, but in Fig. 14(e) is traced out, by greatly reducing the sweep speed, the later history of a similar "first open;" in this case, after a first open had lasted for 12 microsec., there was a second arc which shows clearly on the figure, and this was followed by a second open lasting for 4 microsec. and finally by closure which is recognized by the oscillation.

Even more varied phenomena are exhibited by the 200-volt patterns of Fig. 15. In Figs. 15(a), 15(b), 15(d), 15(f), and 15(h) are shown closures, respectively, after a first, second, third, fourth, and fifth arc; and opens after a second, third, fourth, and fifth arc appear in Figs. 15(c), 15(e), 15(g), and 15(i). Subsequent history of a second open, such as that of Fig. 15(c), is shown in the low sweep speed photograph of Fig. 15(j); on this figure the second arc appears as a small spot, followed by a second open lasting about 15 microsec., then in succession by another spot representing the third arc, a third open lasting 9 microsec., a fourth arc and a fifth arc appearing as spots, a fifth open of 8 microsec., and a final closure of which the first 7 microsec. can be seen.

Except for the absence of a first open the series of traces of Figs. 15(a)–15(i) seems to be complete in exhibiting every possible behavior. It is interesting that there is no example of a first open among a total of 285 clear-cut photographs of closures at 200 volts under the same experimental conditions. In fact, on examination of Fig. 11, one sees that, under the different circuit

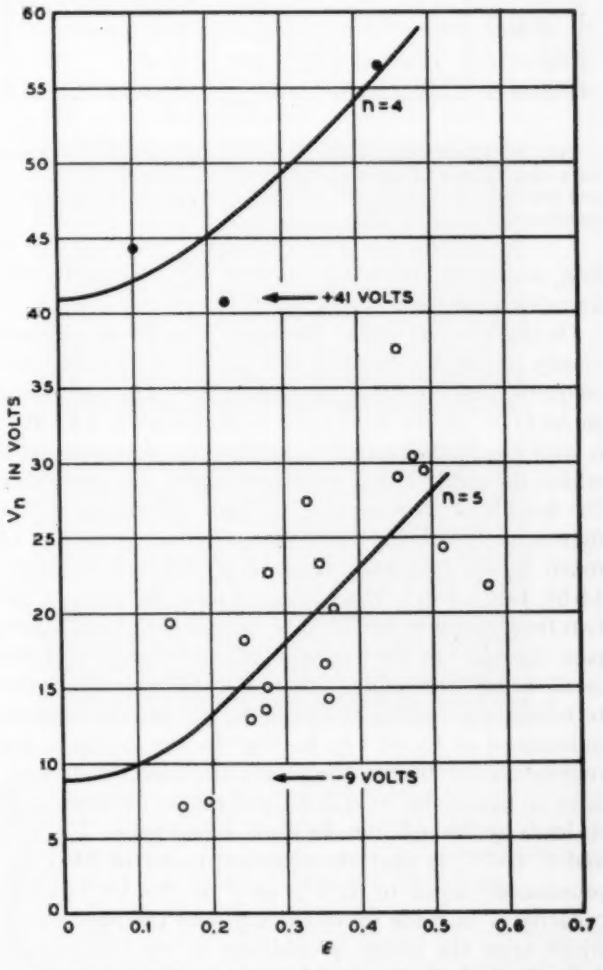


FIG. 20. Variation of open-circuit potential with duration, $\epsilon\pi(LC)^{\frac{1}{2}}$ sec., of a preceding transient closure. The abscissas are values of the proper fraction ϵ in this expression. The solid curves are the result of calculation on the assumption that no energy is dissipated during a transient closure.

conditions of this figure, there was not a single case of a first open for a voltage as high as 200, even from among the total of 7251 different experiments represented by that figure. For a voltage as high as 200, a second discharge always follows immediately after the

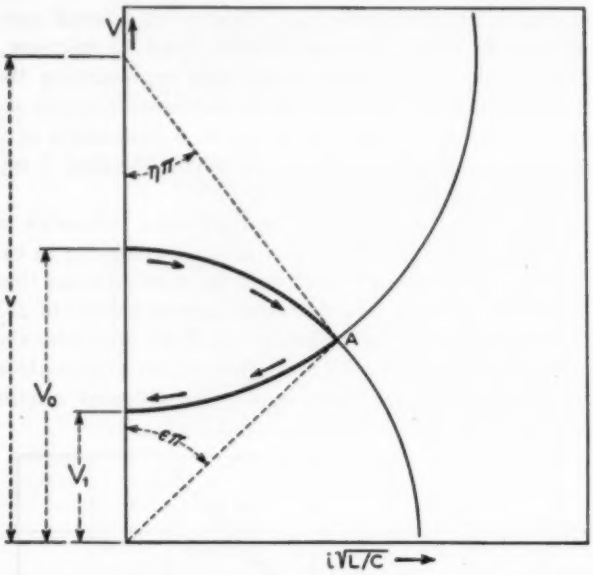


FIG. 21. The relation between current and condenser voltage, for a closure from an initial voltage V_0 burning out at the point A and followed by a discharge at v volts lasting until the current becomes zero.

first, unless of course the contacts are welded by the first discharge.

Oscillations of two different frequencies appear clearly in the photographs of Figs. 14 and 15. The low frequency oscillation upon closure has a period equal to $2\pi(LC)^{\frac{1}{2}}$, for the known circuit constants $L=1\times10^{-6}$ h and $C=20{,}000\times10^{-12}$ f, within the precision with which the sweep speed is known; this period is double the length of each one of the arcs appearing in the figures, as the theory requires. A second oscillation of much higher frequency occurs on open circuit (Figs. 14(b), 14(d), 15(c), 15(e), 15(g), 15(i)), the ratio of the two frequencies being 23. The open circuit oscillations pass through the very small condenser made up of the oscilloscope leads and plates, and its frequency is determined primarily by this capcity and the experimental inductance of 1×10^{-6} h. In Fig. 16 the contacts are represented by arrows, the principal circuit which oscillates on closure by solid lines, and the oscilloscope with its leads by dashed lines. In these experiments, $L'<<L$ and $C'<<C$, so that the observed factor of 23 is approximately equal to $(C/C')^{\frac{1}{2}}$ or C' is 38×10^{-12} f. The capacity C' is made up of the input to the mechanism which trips the sweep in addition to the oscilloscope deflecting plates, and this calculated value seems reasonable.

During each of the arcs recorded on the traces of Figs. 14 and 15 the oscilloscopic record indicates considerable erratic variation of potential, but during the course of each arc there is no systematic variation which is not probably attributable to instrumental imperfection. The arc current, however, rises from zero to a maximum and falls back to zero during each arc, the maximum being of the order of 10 amp. for the first arc at 100 volts and 20 amp. for the first arc at 200 volts. The lack of any systematic change of the voltage over the wide current range is notable, and consistent with the observations of Holm.[5] We believe that the voltage being independent of current must be interpreted to mean that these short arcs have no positive column.

C. Transient Closures

The various open circuit voltages of Figs. 14 and 15 fit only moderately well upon the plot of Fig. 11. A good fit is not to be expected because the circuit inductance and the circuit capacity were widely different when these patterns were obtained from the values when the traces were made which gave the points of Fig. 11, and therefore the parameter f was different for the two sets of data. The open circuit voltages of Figs. 14 and 15 can be expected to fit upon a plot like that of Fig. 11 with a slightly different value of f.

A test of this presumption has, in addition to establishing its validity, brought into prominence another phenomenon of contact closure which up to this time had been observed only imperfectly, and has in so doing accounted for some small part of the considerable erratic variation of the experimental points of Fig. 11. An examination has been made of 369 photographs of contact closures under the experimental conditions of the patterns of Fig. 15. Of these, 285 turn out to be free from gross faults which make them unsuitable for study, and there are on these photographs 247 open circuits. Most of these resemble the patterns of Fig. 15, but 36 exhibit a phenomenon not heretofore reported. Sample records of this phenomenon in its simplest form are reproduced as Figs. 17(a) and 17(b), and it appears with some added features in Figs. 17(c), (d), (e), and (f).

On the trace of Fig. 17(a) one sees second and third arcs with the latter ending in a closure, as in Fig. 15(d), but this closure lasts only about 0.1 microsec. with the contacts opening to a fourth arc, and the behavior is "normal" thereafter. Figure 17(b) is similar, with the transient closure occurring just after the fourth arc rather than after the third. The interpretation of these patterns is obvious; when the first closure occurs the metal bridge is so narrow that it is rather promptly burned out by the flow of current through it.[6] Figures

[6] An alternative explanation might conceivably attribute the reopening of the contacts to chatter, but this is certainly not correct because of the exceedingly brief time of the closure. A number of experiments have been made to find out how severe chatter can be. It is easy to produce chatter, in fact moderately difficult to eliminate it. One cannot, however, produce chatter in which the contacts open indefinitely soon after closure. By adjustment of amplitude of the vibrating electrode, and contact pressure

17(c) and 17(d) resemble, respectively, 17(a) and 17(b) in the occurrences of transient closures, but in 17(c) and 17(d) each closure is immediately preceded by an open circuit of brief duration.

Occasionally more than one transient closure appears on the same record, and the most elaborate pattern which has been found is reproduced as Fig. 17(e). There are *four* transient closures on this trace and the record ends with a *seventh* open! This is striking because under the experimental conditions of these photographs there can be, in general, only five arcs, the condenser energy being almost dissipated at the end of the fifth. When, however, the current flows through a solid metallic path for part of one or more of the half-oscillations, the rate of dissipation of energy is very greatly reduced during this part of the time. Thus, an open circuit occurring subsequent to a transient closure must have a potential greater in absolute magnitude than it would have had if the closure had not occurred.

Probably many of the open circuits which are plotted in Fig. 11 were preceded by transient closures and each of these closures must have increased the measured open circuit voltage and thus increased the scattering of the experimental points. The traces at 200 volts, examples of which appear in Figs. 15 and 17, are suitable for study in this connection. Among these there are 36 opens which follow transient closures and 211 others which are not preceded by closures. The opens of the latter group occur at significantly lower voltages than those of the former of the same order. Voltages of 2nd, 3rd, 4th, and 5th opens, which are *not* preceded by transient closures and which can be measured with fair precision, are plotted on Fig. 19, the ordinates being the number of times each potential was recorded. The arrows on this figure indicate the open circuit voltages calculated from Eq. (5) for $v=15$ volts and $f=0.92$. These arrows fit the experimental data within the uncertainty involved, and about as well as the data can be fitted if we hold to $v=15$ which was obtained from Fig. 11. An equally good or better fit would be given by substituting $v=17$, $f=0.95$ into Eq. (5). Although one source of systematic error was eliminated by omitting those open circuits which were preceded by transient closures, the scatter of the points is still so bad that v cannot be determined within 3 or 4 volts.

The experimental value $f=0.92$ corresponds to $R=0.51$ ohm for the values of inductance and capacity of these experiments. It is interesting to compare this with values of high frequency resistance obtained in other experiments. The trace of Fig. 13(c) gives $R=0.40$ ohm at a frequency of 0.7×10^6 cycles/sec., the data of Fig. 19 give 0.51 ohm at 1.1×10^6, and those of Fig. 11 give 0.77 ohm at 20×10^6. The direct current resistance

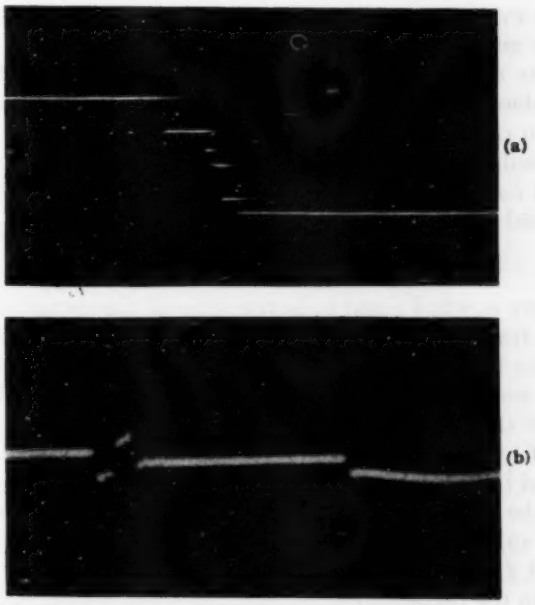


FIG. 22. Potential across contacts on closure from low voltages. (a) Fastax camera trace of closure from 16 volts with intermediate opens at 11.3, 9.0, 6.8, 6.5, and 2.0 volts. (b) Detailed record at 2-microsec. sweep time of closure from 10 volts, showing transient closure followed by a discharge at about 15 volts and ending in an open at 6 volts, with final closure after 0.9 microsec.

and velocity, it has been possible to obtain chatter in which the first open occurs about 7 microsec. after closure, but it has not been possible to obtain a shorter time interval than this.

This observation of a minimum possible interval agrees with unpublished experiments by Dr. J. J. Lander and with an unpublished theoretical investigation by Dr. R. D. Mindlin. For the impact of a particular cantilever bar against a stop, Mindlin was able to calculate that the first possible open should come at 400 microsec. after impact, and experiments upon this bar gave a first open after 250 microsec.; the numerical values of the constants entering into the calculation were imperfectly known, and these two values were considered by Mindlin to be in satisfactory agreement. The vibrating system in the present experiments is very much lighter than that upon which Mindlin and Lander worked, and it seems reasonable that for this system the first possible open should be at 7 microsec. which is the smallest value we have observed.

Some oscilloscopic traces of voltage across contacts on closure, which were obtained in the course of the studies of chatter, are interesting enough to deserve reproduction here, even though they are off the main line of the present investigation. The pattern of Fig. 18(a), obtained without disconnecting the voltage source at closure, as is usually done, shows a first mechanical open at about 7 microsec., followed by reclosure at 36 microsec., a second open at 47 microsec. and a second reclosure at 60 microsec. The traces of Figs. 18(b) and 18(c) show details of the first open at 7 microsec., with the voltage source disconnected at closure in the usual fashion. The oscillation for the first 7 microsec. of the trace of Fig. 18(b), has a frequency of 2.6×10^6 which is appropriate to a closure. At 7 microsec. occurs the first mechanical opening, following which the potential across the contacts oscillates at a much higher frequency. This frequency can be measured to be 12×10^6 on Fig. 18(c), the pattern of which was photographed at a higher sweep speed with the initiation of the sweep delayed by an appropriate network. The most interesting feature of these patterns is the increase in the amplitude of the closed circuit oscillation as the time of contact opening is approached; this is due solely to increased resistance at the contacts, making the voltage greater even when the amplitude of the oscillating current is necessarily decreasing. For the pattern of Fig. 18(d) the contacts were adjusted to be free from chatter but to be very close to conditions which cause chatter. One can see here the increase in the amplitude of the closed circuit oscillation as the 7-microsec. point is approached, and then the subsidence of the oscillation as the contacts are finally pressed together instead of separating.

must have varied slightly in these experiments but in the reverse order.

It is stated above that those open circuits which are preceded by transient closures occur at significantly higher voltages than those plotted on Fig. 19. The differences are shown by the plots of Fig. 20; here the ordinates are voltages of those opens which were preceded by one, and only one, transient closure, and the abscissas are the fractions of the preceding arc time taken up by this closure. The solid curves were calculated by Dr. J. M. Richardson on the assumption that the energy loss is strictly zero during the time of each transient closure, which is a very good approximation. The consistent deviation of the measured voltages above the values for no transient closure, marked by the arrows, proves that the selection of the data of Fig. 19 eliminated one source of systematic error which was no doubt present in the plot of Fig. 11, and the agreement of the points with the theoretical curves on Fig. 20 indicates that we understand the phenomenon.

It is clear that transient closures are not the only cause of the scattering of the experimental points of Fig. 11. There is considerable actual variation in the arc voltage and this shows up in variability of subsequent open circuit voltages. There is a variation of arc voltage not only between one discharge and the next, but even during the course of a single discharge. This one might suspect from the traces of Fig. 15 and 17. It is confirmed by detailed and as yet unpublished studies of arc voltages made by Mr. H. J. Juretschke.

In calculating the curves of Fig. 20, Richardson assumed, as a simplification of the algebra, that the circuit contained no resistance, that is, $f=1$. He then obtained

$$V_{n+1}=-V_n[(1-2\alpha_n\cos\pi\epsilon_{n+1}+\alpha_n^2)^{\frac{1}{2}}-\alpha_n], \quad (6)$$

where $\alpha_n=v/V_n$ and ϵ_n is the ratio of the duration of the transient closure in the n-th arc, assumed to occur at its beginning, to the total length $\pi(LC)^{\frac{1}{2}}$ which the arc would have had if there had been no closure. (For $\epsilon=0$ this formula becomes equivalent to Eq. (5) for $f=1$.) The solid curves of Fig. 20 represent V_n calculated from Eq. (5) for $v=15$ and $f=0.92$ (the constants of the arrows of Fig. 19) plus the difference between V_n calculated from Eq. (6) and V_n calculated for $\epsilon=0, f=1$.

On the traces of Fig. 17 each transient closure is followed by an arc, and this observation is true of each of the 36 transient closures which were recorded in this series of observations. It appears that an arc must always occur after a contact has been burned out by current flowing through it. We shall see below that this is true even when the initial circuit potential is below the discharge voltage v.

It is interesting that each of the transient closures of Fig. 17 was burned out before the end of the first half-cycle of the oscillating current flowing through it. In fact we have no record of a burn-out occurring after the end of the first half-cycle; if a closure lasts for a half-cycle it is permanent. Most burn-outs occur during the first half of the half-cycle, while the current is increasing, and only rarely, as, for example, in the fifth arc of the trace of Fig. 17(e), during the second half after the current has passed its maximum.

The phenomenon of a discharge following each transient closure is always essentially the same, even when the initial potential V_0 is below v, the arc potential. In this case, when burn-out occurs the energy stored in the circuit inductance raises the potential across the contacts up to the arc potential to conclude the half-cycle of the oscillating current with an arc. The situation is shown most clearly by the plot of Fig. 21, due to Richardson, giving the relation between voltage across the condenser and current in the circuit. The circle about the origin corresponds to a closure, and that about the point $(0, v)$ to an arc starting from a burn-out at an arbitrary point A on the closure circle and ending at the open circuit potential V_1. (Equation (6) is determined at once from Fig. 21.) If we set the scale by $v=15$ volts, then this sketch represents a transient closure from $V_0=9$ volts lasting for $\epsilon\pi(LC)^{\frac{1}{2}}=0.82(LC)^{\frac{1}{2}}$ sec. and ending in an arc for $\eta\pi(LC)^{\frac{1}{2}}=0.60(LC)^{\frac{1}{2}}$ sec. with a final open at $V_1=4$ volts. The total time of the transient closure plus the arc is $1.42(LC)^{\frac{1}{2}}$ sec. which is, in this case, a little less than half of the normal arc time. There is no qualitative distinction between an arc following a transient closure for $V_0<v$ and one following a closure for $V_0>v$.

Transient closures are frequent at low values of V_0. Of a group of 49 photographs of closures at $V_0=12$ volts only 20 are without open circuits, the other 29 exhibiting one or more opens at voltages which are widely distributed. There are 15 single opens, 10 double, 3 triple, and 1 quintuple. A photograph of the last of these is reproduced as Fig. 22(a).

Figure 22(b) is a trace[d] of a closure from 10 volts with a total sweep time on this enlarged reproduction of 2×10^{-6} sec. One cannot read off reliable values of voltages and times from this trace because of failure of the oscilloscope to respond sufficiently rapidly, but the figure indicates a transient closure lasting about 0.07×10^{-6} sec. followed by an arc at a voltage of the order of 15 lasting for about the same time and ending in an open circuit at 6 volts. Final closure occurs about 0.9×10^{-6} sec. later. We believe that each of the opens of Fig. 22(a) occurred after a transient closure ending in a discharge at a potential of the order of 15 volts, as in the record of Fig. 22(b). The details of the transient closures and succeeding arcs cannot show up on

[d] The trace of Fig. 22(b) was obtained in somewhat indirect fashion by tripping the oscilloscope sweep on a normal discharge which carried V_0 from 21 volts to 10 volts, using a considerable delay in the sweep circuit so that out of many photographs we could, now and then by chance, catch the closure from 10 volts. This procedure was necessary because we cannot start the sweep sufficiently rapidly to record the beginning of a closure with adequate resolution.

Fig. 22(a) because of the low equivalent sweep speed; in fact, even the "final" open circuit of Fig. 22(b) is of such brief duration that it could scarcely have been detected on such a trace as that of Fig. 22(a).

Other examples of low voltage closures are shown in Fig. 23. These traces are Fastax camera records of closures from 21 volts. In each case the first feature is an open at about 11 volts and this is followed, except in Fig. 23(a) by open circuits which are the result of transient closures ending in discharges. It is of interest to note that open circuits after an initial voltage of $V_0=11$ are widely scattered in value between 0 and 11 volts, but that an open immediately after $V_0=21$ volts tends to occur near 11 volts. The former opens are presumed to be the results of arcs which were *initiated* by transient closures whereas most of the latter were the results of arcs without initial closure; there cannot, in fact, be an arc for $V_0<15$ volts without initial closure, but if V_0 is much greater than 15 volts initial closure is not required.

There are, however, in the traces of Fig. 17, four examples of arcs following fairly high voltage transient opens, and each of these is initiated by a transient closure. In all of the oscilloscopic records at 200 volts there are 10 examples of the ends of transient opens. Two of these opens end in permanent closures without arcs, two end in arcs without detectable closures, and 6, of which 4 are reproduced in Fig. 17, end in transient closures followed of course by arcs. *These records must be taken as evidence that in many cases an arc is initiated by a closure the burning out of which furnishes sufficient metal vapor for the discharge.*[6] All of the transient opens of Fig. 17 are necessarily short; otherwise their ends could not be observed. It is to be presumed that an arc does not, in general, start from a transient closure when the preceding open has not been short. We have in mind that adsorbed gas is necessary in initiating an arc before electrodes touch, and that at the end of a very short transient open this gas supply has not been replenished after removal in the preceding arc and the arc cannot therefore start without mechanical contact; but, when the electrode surfaces are well covered by gas an arc is struck before the electrodes touch. A theory of how this might happen was suggested by us earlier,[3] but seems to be inadequate to account for more recent tests described in the following section.

D. Arrested and Suppressed Discharges

The oscilloscopic traces of Figs. 12(b) and 13(c) represent closures without the discharge at about 15 volts which has been discussed above in such detail. Whether or not a 15-volt arc takes place on closure is determined by a number of factors which have been investigated in empirical fashion. Since erosion on make is presumably due entirely to the arc it is of the greatest importance to understand how the arc can be prevented. We shall give below only a few of the observations which have been made, some of them being at the present time too confused for brief presentation.

In many closures, on which there is no arc at the usual

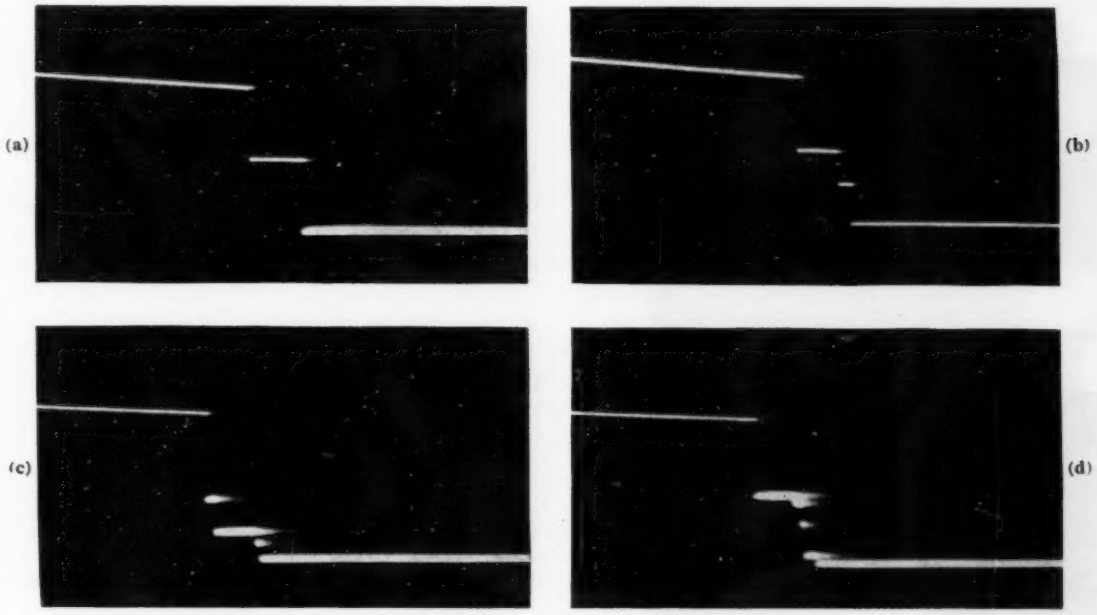


FIG. 23. Fastax camera trace of potential across platinum contacts showing closures at low potentials. Applied potential difference 21 volts. Each trace shows an open circuit at about 11 volts which is the result of a normal discharge, but the phenomena of final closure are widely different in the four cases, each except the first exhibiting a number of transient closures before final contact is established.

[6] This is the mechanism to which Holm (see reference 5, p. 280) ascribes the starting of a discharge on make.

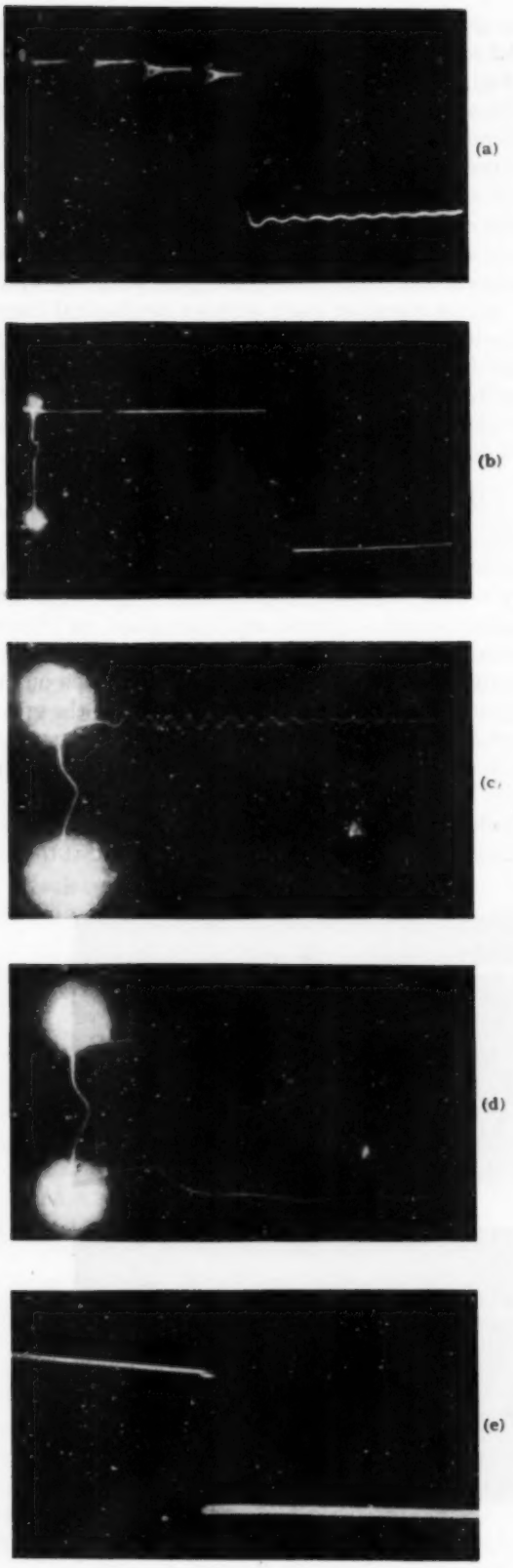


FIG. 24. Potential across contacts, showing arrested discharges. (a) Four successive arrested discharges before final closure. Sweep time 30 microsec. (b) Two arrested discharges, followed by a normal arc ending in an open. (c) An arrested discharge recorded at high sweep speed to show open circuit oscillations. Sweep time 1 microsec. (d) Closure recorded at high sweep speed, to show absence of normal discharge—presumably following an arrested discharge. (e) Fastax camera record of an arrested discharge.

15 volts, an oscilloscopic record shows that potential across the contacts decreases abruptly by a very few volts and then remains steady for a considerable time; the discharge obviously starts but is arrested by circumstances unfavorable to its continuance. Frequently there are many such arrested discharges before closure, a record of four such successive arrests being reproduced in Fig. 24(a). After arrested discharges contacts may close without a normal 15-volt arc, as in the experiment of Fig. 24(a), or with a normal arc as in that of Fig. 24(b).[e] It is our present view that a discharge may perhaps start on every closure if the initial voltage is well above 15. Under conditions favorable to it the voltage drops rapidly to about 15 to complete the arc with the resulting erosion of metal from the positive electrode, but under less favorable conditions the discharge may be arrested when the condenser has lost only a small fraction of its charge. After such an arrest one or more new discharges may be started and the final one may or may not develop into the low voltage arc which causes damage to the anode. The fact that closure *can* occur without this arc opens up the possibility that discovery of the way in which it is initiated may suggest a means for its prevention, and thus practically eliminate contact erosion on make.

For platinum electrodes operating in air at voltages above 50, a 15-volt arc will occur on each closure if the circuit inductance (L in Fig. 16) is below 1 microhenry, but will rarely occur if the inductance is above 3 microhenries.[f] (Special circumstances modifying this statement will be described.) The limiting value of inductance depends upon capacity and upon voltage; high voltages and high capacities favor the arc but the major factor is circuit inductance and the limiting value of inductance varies only moderately for wide changes of capacity and potential.

Very great changes in the limiting value of circuit inductance can, however, be brought about by changes in the surfaces of the electrodes caused by impurities in the atmosphere. Exposure to a low concentration of certain organic vapors can, for example, so change the electrode surfaces that a 15-volt arc of normal type will occur upon closure at 50 volts for a circuit inductance as high as 1000 microhenries, and a somewhat modified type of arc discharge will occur for indefinitely high inductances, up to at least 12 henries. The change is not brought about by simply exposing the contacts

[e] The high frequency oscillation, which occurs on an open circuit, takes place also on an arrested discharge. One can see this on the arrests of Figs. 24(a) and (b), but more clearly on Fig. 24(c) for which the sweep speed was increased to give greater resolution.

[f] In earlier sections it was pointed out that pit volume, and doubtless metal transfer, are greatly reduced by inductance. It is now clear that the major factor in this reduction is the effect of the inductance in preventing an arc.

to the vapor; they must discharge a condenser repeatedly in the vapor, about 10^4 times in one particular experiment, before the arcs occur. Furthermore, after this treatment the vapor is no longer essential; an arc will continue to occur on closures taking place in the absence of vapor until, after about 10^4 closures in one particular experiment, the changes produced by the vapor have been destroyed. Then, again, an arc at 15 volts will occur upon closure only when the circuit inductance is not above 3 microhenries. (If the contacts are not operated, or are operated without the electrical circuit, their active condition is maintained indefinitely.) These observations and changes seem to be very similar for electrodes of platinum, palladium, gold, and silver. It is appropriate to speak of the electrode surfaces as being in an active condition when they will produce an arc discharge through a circuit inductance as high as 1000 microhenries, and in an inactive condition when they will not produce such a discharge. In our experiments all of the compounds which have produced the active condition have been unsaturated compounds, and most unsaturated compounds tried have been found effective. Oscilloscopic records of closures of platinum electrodes in the active condition are reproduced as Fig. 25. Corresponding records with electrodes in the inactive condition would show only the initial vertical line and the horizontal line representing final closure.

According to our present view the active condition of electrodes, which permits an arc to be established through a large circuit inductance, consists of a surface layer of carbon or a carbonaceous compound which is capable of holding a high surface density of adsorbed gas and of giving an insulating surface film. This view is supported by the observation that a make arc is always struck between carbon electrodes regardless of the size of the inductance; a carbon electrode is never "inactive" in the sense that electrodes of platinum, palladium or gold can be inactive. Experiments in which one electrode is carbon and the other platinum are informative. If the carbon electrode is negative a make arc is always struck but if the platinum electrode is negative an arc through high inductance is struck only when the electrodes have been activated by a procedure which would be adequate to activate two electrodes of platinum. This result seems to require that, whatever the activation process may be, the active condition of electrode surfaces is a quality associated with the cathode surface. It is, furthermore, a quality that is always possessed by carbon, which is in accord with the observation that platinum surfaces are activated by long continued operation in an atmosphere containing any one of certain organic vapors.

5. FIRST ATTEMPT AT A THEORY OF THE DISCHARGE

The 15-volt discharge,[g] which is responsible for the erosion of electrical contacts on make and with which

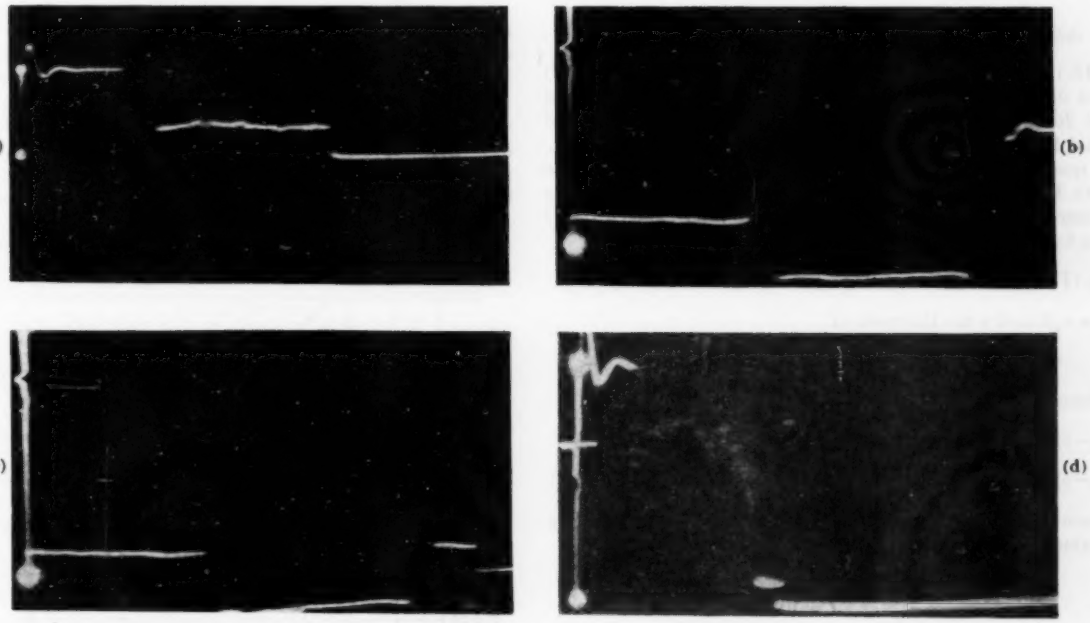


FIG. 25. Potential across contacts on closure, in the "active condition." $10{,}000\times10^{-12}$ f, 1000×10^{-6} h, sweep time 30 microsec. (a) 50 volts. An arrested discharge ending in an open circuit, followed by about four arrested discharges following each other in rapid succession and ending in a 15 volt arc and final closure. (b) 150 volts. First and second 15-volt arcs ending in a second open. (c) 150 volts. First, second, and third 15-volt arcs, ending in closure. (d) 150 volts. An arrested discharge ending in an open circuit, followed by a large number of arrested discharges in rapid succession ending in a 15-volt arc and final closure.

[g] The expression "15-volt discharge" is intended to be a means of identification only. Juretschke has shown that the potential of this discharge is different for different electrode materials. In particular, the potential is considerably higher for carbon than for any metal which has been tested.

most of this paper is concerned, is presumed to be a true arc in metal vapor. This arc can be initiated by field emission of electrons when approaching electrodes have come close enough together to give a sufficiently high field at the cathode; the arc follows also after a transient closure, when the first point of contact of two electrodes happens to be so small that it is burned out by the rush of current through it. In the latter case one understands how the arc can be established in the metal vapor resulting from the burn-out, but when the first event leading to the arc is the appearance of field emission electrons one needs to consider carefully just where the metal vapor comes from.

We have postulated the ionization of adsorbed gas molecules on the anode by the electrons and an increase in the field at the cathode caused by the approach of ions from the anode, to give rise to further electrons.[3] The cumulative rise of current by this process gives an electron avalanche[h] which very soon vaporizes metal from the anode in significant amounts. Whether or not an arc is established in the metal vapor depends upon how efficient each ion is in drawing out electrons from the cathode, upon the available supply of gas on the anode, and upon the rapidity with which this supply is converted into ions. The "activity" of a cathode surface can be accounted for by the postulate that an active surface is covered by an insulating film, the outer surface of which becomes charged by positive ions whose efficiency in drawing out electrons is then greatly increased.[7] The various aspects of this theory have at present relatively little experimental confirmation and we view the theory only as a convenient framework around which future work can be planned.

A large part of the experimental work reported in this paper was very ably carried out by Mr. J. L. Smith to whom we are greatly indebted.

[h] We are sure that the progressive build-up of current by this means is not the entire phenomenon of the arc discharge, as postulated earlier (see reference 3), because the final voltage is always about 15 whatever the value of the initial voltage, even when the latter is very many times greater than 15 and the impressed field at the cathode must therefore drop abruptly by a very large factor without ending the discharge. This fault of the original theory, which was recognized by us early, was also pointed out to us strongly by Dr. J. Slepian.

[7] This is related to the "Spritzentladung" discovered by A. Güntherschulze and his collaborators and to the "Malter effect."—M. J. Druyvesteyn and F. M. Penning, Rev. Mod. Phys. 12, 87 (1940).

APPENDIX A. EFFECT OF STEADY VALUE CURRENT UPON EROSION

D. P. Ling
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

THE circuit of experimental contacts and of a capacity which is discharged across them is represented by the solid lines of Fig. 26. The dotted lines represent the source of power, of potential V_0, and its associated wiring.

The quantity of interest is i, the current discharging across the contacts. Under conditions which correspond to the instantaneous establishment of the potential v across the contacts, the current i is given by the following expression:

$$i=(V_0-v)/R-[(V_0-v)/R+\xi]e^{\gamma t}+e^{\alpha t}(\xi\cos\beta t+\eta\sin\beta t),$$

where $\alpha\pm j\beta$ and γ are the roots of

$$r^3+(R/L')r^2+[(L+L')/CLL']r+R/CLL'=0,$$

and where

$$\xi=(V_0-v)[2\alpha(R+L\gamma)-L\gamma^2]/LR[(\gamma-\alpha)^2+\beta^2]$$

$$\eta=(V_0-v)[(R+L\gamma)(\beta^2+\gamma^2-\alpha^2)+L\gamma^2(\alpha-\gamma)]/LR\beta[(\gamma-\alpha)^2+\beta^2].$$

In certain of the experiments reported in this paper the following fixed values are appropriate:

$V_0=50$ volts $\qquad L=0.5\times10^{-7}$ henry
$v=15$ volts $\qquad C=10^{-9}$ farad

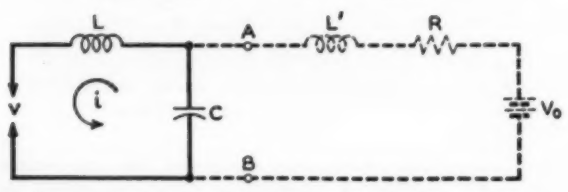


FIG. 26.

and our concern is with the performance of the system as R and L' are allowed to vary.

Upon activation of the circuit, the current rises to begin a series of oscillations. Under appropriate conditions on R and L', i may remain positive throughout these oscillations. In this case V_0, the external source of power, will contribute a good deal of energy to the arc by means of the steady-state current $(V_0-v)/R$ which eventually establishes itself. Alternatively, the current may drop negative on its first downward swing, in which event the arc disappears, and the charge which has passed in the course of the discharge time t will have been only slightly larger than $2C(V_0-v)$, and thus be practically unaffected by the potential source. This is the situation that would account for the experi-

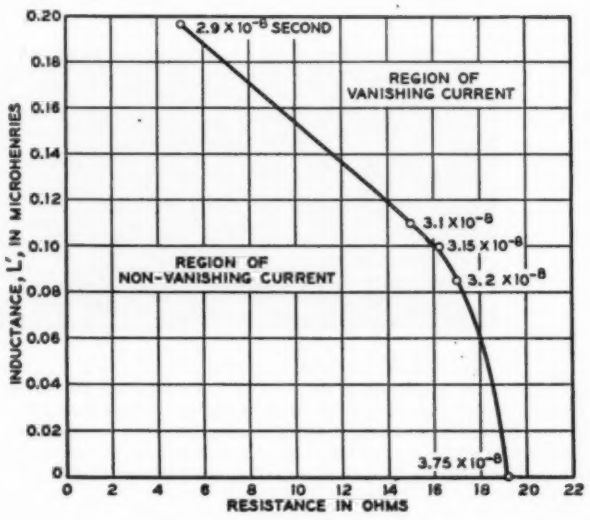


FIG. 27.

mental observation that the amount of transferred metal is independent of the steady value current.

A calculation was made on the basis of the formulas just given to determine the relation between L' and R which separates the vanishing from the non-vanishing case. The results of the computation are presented in Fig. 27. For any R, the L' of the curve is that for which the current just touches the axis $i=0$ on its first downward swing. Points above the curve correspond to pairs of values of R and L' for which the current will vanish, those below the curve to those for which it will remain positive. The circles represent the computed points, and the number beside each is the time t from activation of the circuit to the vanishing of the current. From this curve one sees that the power source will contribute little to the arc energy if the resistance in series with the source is as large as 20 ohms, corresponding to a steady value current of $(50-15)/20=1.75$ amp. In this paper was reported the experimental observation that, for circuit constants approximating those of the calculation, the source of power had no effect upon erosion unless the steady value current considerably exceeded 1 amp.

In the case of telephone relays, L and C of Fig. 26 represent the lumped equivalents of the distributed inductance and capacity of the lead wires. These leads exceed 2 feet in length, but rarely exceed 100 feet. Their inductance will run about 0.3×10^{-6} h per foot, their capacity 20×10^{-12} f per foot, though this latter figure is subject to a considerable variation with humidity. We may conclude, however, that the value of C used in the calculation is of the order of magnitude of corresponding capacities in telephone relay circuits, and that the value of L, although smaller, does not differ by many orders of magnitude. On the other hand, for telephone relays the minimum value of R is 100 ohms and of L' is 0.3 h. (R may run to over 400 ohms, L' up to 10 h.) Consequently, it is safe to conclude on the basis of Fig. 27 that *in telephone relay circuits* the current of an electrical discharge occurring on closure reaches the zero axis and the discharge receives therefore only a negligible direct contribution of energy from the potential source.

A minor complication is presented by the presence in actual circuits of this sort of a protective circuit bridging the lines from A to B (Fig. 26), and designed to protect the contacts at break. This consists of a resistance and capacity in series. The resistance is 100 ohms or more, the capacity 0.1 to 1×10^{-6} f. With a condenser of this size, this circuit behaves much as would a second external power source, and the validity of our argument is unaffected.

The Scattering Cross Section of Spheres for Electromagnetic Waves

L. BRILLOUIN
*Cruft Laboratory, Harvard University, Cambridge, Massachusetts**
(Received May 11, 1949)

The problem discussed in this paper is that of the scattering of electromagnetic waves by a spherical obstacle. The classical theory is well known. Stratton's "Electromagnetic Theory," for instance, contains a very good summary of that discussion. The computation of the scattering cross section, according to this theory, leads to some difficulties. In the case of large spheres, where geometrical optics should apply, the rigorous theory yields a scattering cross section equal to twice the actual cross section of the sphere!

The discussion presented in this paper explains this strange result and shows the role played by the shadow and by the diffraction fringes surrounding the shadow. A reasonable system of approximations yields the well known "Babinet's principle." The physical interpretation is of such general character that it must certainly apply to a variety of similar problems in acoustics or wave mechanics. The spherical shape of the obstacle is essential in the present discussion, but similar results would certainly be found for other shapes of the obstacle. The case of a circular cylinder, investigated by E. B. Moullin and L. G. Reynolds, also leads to a total cross section twice as large as the actual cross section, for large cylinders. The explanation is the same as for the spheres. Experiments very carefully designed by Dr. Sinclair and Professor V. K. LaMer gave a very precise check of the theoretical predictions.

I. INTRODUCTION

THE investigation is concerned with the analysis of inconsistencies and common misinterpretations found in several standard books and research papers dealing with the problem of diffraction of light by spherical obstacles. It is the object of this investigation to clarify these inconsistencies and to provide a physical interpretation of theoretical results that is in accord with experiments.

The problem of the scattering of light by small spherical obstacles is an old one in theoretical physics, and goes back to Rayleigh, Debye, and Mie, whose papers on the subject became classics. The theory is known to give very good results as far as angular distribution of secondary radiation is concerned. It checks well with experiments, provided scattering particles are true spheres of well-defined radius. However, there appears to be a serious difficulty about the absolute value of the total amount of energy loss per second (energy lost by the beam and partly absorbed in the sphere, partly reradiated in all directions).

V. K. LaMer[1] pointed out that the theoretical value for the energy scattered per second by the sphere as computed, for example, by using the formulas given by Stratton[2] is about twice as great as the experimental one. LaMer and Sinclair suggested that the customary calculation, according to a pattern proposed by Mie, ought to be in error at the point where the superposition of two independent beams of polarized light is used to obtain the formulas for natural light. It is true that many authors, including Rayleigh, made errors at this stage of the computation.[3]

It is demonstrated that another reason for this discrepancy is far more fundamental, and it is shown that if one uses the usual method of computing the energy scattered per second by the sphere for *polarized incident light* the scattering cross section area Q_s for a reflecting sphere of large radius a is $Q_s=2\pi a^2$, which is twice the value given by geometric optics. Moreover, this difficulty is not connected with the shape of the body, as can be seen from certain British computations[4,5] which treat the problem of light scattering by a solid cylinder of radius a and obtain an apparent scattering cross section, which for a wave-length that is small compared with the radius of the cylinders tends to a value $4a$ instead of $2a$ (per unit length of the cylinder).

The true explanation of such discrepancies is found to lie in the proper interpretation of the use of Poynting's flux theorems and in the special role played by the shadow of the body. In order to obtain the correct value for the scattering cross section, it is necessary to take into account the electromagnetic fields in the shadow region and in the scattered radiation.

The difficulties connected with the role of the shadow do not arise in the problem of radiation pressure, since all the energy radiated at small angles of deflection does not contribute to the radiation pressure. This explains why Debye obtains the correct result for the radiation pressure on a metallic sphere of large radius. The problem is a fundamental one in wave theory, and similar considerations apply to acoustics or wave mechanics.

The following section is devoted to a brief summary

* Now with the International Business Machines Corp., New York, New York.

[1] Victor K. LaMer, "Verification of Mie Theory, Calculations and Measurements of Light Scattering by Dielectric Spherical Particles," Progress Report, Div. 10, OEMsr-148, Service Project CWS-1, OSRD 1857, Sept. 29, 1943, AMP-202-M1.

[2] Julius Adams Stratton, *Electromagnetic Theory*, (McGraw-Hill Book Company, Inc., New York, London, 1941), p. 568.

[3] W. S. Stiles, "On the Intensity of Scattered Light from an Unpolarized Beam," The London, Edinburgh and Dublin Philosophical Magazine and J. Science (London) **7**, 204–205 (1929), Sec. XXIII.

[4] E. B. Moullin and L. G. Reynolds, "The Problem of a Plane Wave Incident on a Circular Cylinder Parallel to the Electric Vector," Report M.397, OSRD WA-933-1, Admiralty Signal Establishment (Great Britain), December 1941, AMP-203-M1.

[5] E. B. Moullin and L. G. Reynolds, "Expression for the Echoing Power of Cylinder," Appendix to Report M.397 (attached to Report M.397), OSRD WA-933-2, Admiralty Signal Establishment (Great Britain), December 1941, AMP-203-M1. Also: C. H. Papas "Diffraction by a cylindrical obstacle" Cruft Lab. Report No. 38 ONR June 20, 1949.

of the optical properties of spheres and the amounts of light they scatter when the problem is analyzed from the point of view of geometrical optics and elementary physical optics.

II. ELEMENTARY APPROACH TO THE PROBLEM

Geometrical optics is known to give a correct answer whenever the dimensions of all instruments involved are much larger than the wave-length λ. In our problem this means that the radius a of the sphere and the distance R of observation should be considerably larger than λ

$$a>>>\lambda, \quad R>>>\lambda. \tag{1}$$

For the case of a metallic sphere, acting as an ideal mirror, the result is:

a. Scattering cross section = actual section = πa^2.
b. Light scattered with uniform distribution all around the sphere.
c. A perfectly dark shadow inside a cylinder of radius a behind the sphere. (2)

In this shadow, the electric field is exactly zero.

Elementary physical optics is usually based on Huygens' principle. It is known to yield only a very rough first approximation of a more correct theory of wave propagation. This method does not take into account the transverse character of electromagnetic waves and treats longitudinal or transverse waves in exactly the same way, which is bound to be wrong. Furthermore, this theory makes no provision for the exact shape of the boundary nor for specific boundary conditions. In our problem, Huygens' principle yields similar results for a dark circular disk, an absorbing sphere, or a reflecting sphere, and such a complete similarity between widely different problems cannot be taken seriously.

This theory, however, is known to give information of real practical value and the corresponding results are summarized briefly here. The theory is found in all textbooks on optics.[6]

As is well known, the fringe patterns are similar when observed either behind an aperture of a given shape or behind a screen of the same shape. This is the "principle of Babinet."[7] The fringes behind a circular aperture have been known for a long time; their intensity distribution is governed by a J_1 Bessel function, and the first black rings correspond to

$$p=0.61\frac{\lambda}{a}; \quad 1.115\frac{\lambda}{a}; \quad 1.619\frac{\lambda}{a}\cdots \quad \lambda<<a, \tag{3}$$

where a is the radius of the circular aperture, λ the wave-length and p the ratio

$$p=\frac{r}{R}=\sin\theta,$$

r the radius of a dark ring on a screen located at a distance R from the aperture; θ the angular deflection.[8] Similar rings should be observed in the light scattered by a sphere. They occur for small angles θ and constitute conical fringes around the geometrical shadow.[9] In addition, the theory also predicts a white *bright fringe* of interference all *along the z axis*, just in the middle of the geometrical shadow.

Let us define a cone with a small angle θ_s that would include the first rings (3) and the shadow, and choose

$$\theta_s\approx\sin\theta_s=K\frac{\lambda}{a}$$
$$K\approx 1 \text{ to } 3. \tag{4}$$

We can give a rather crude description of the shadow structure in physical optics in the following way (see Figs. 1 and 2).

A plane wave falling upon the sphere should give a sort of conical shadow region S where the intensity of light ought to be extremely small except for a bright fringe along the axis OSR. This shadow would extend up to a distance R,

$$\frac{a}{R}=\theta_s=K\frac{\lambda}{a},$$
$$R=\frac{1}{K}\frac{a^2}{\lambda}. \tag{5}$$

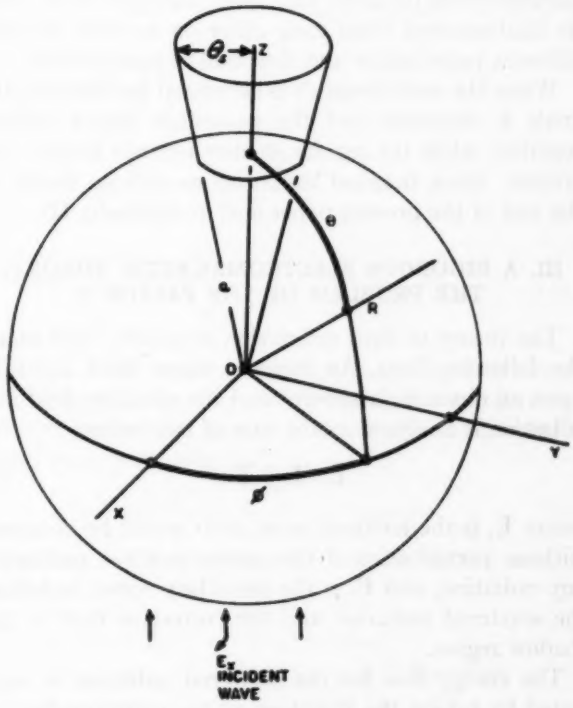


FIG. 1. Coordinate system for shadow structure.

[6] Max Born, *Optik* (Verlag Julius Springer, Berlin, 1933), p. 160.
[7] See reference 6, p. 153.
[8] See reference 6, p. 154, Eq. (10).
[9] See reference 2, p. 284, Eq. (86).

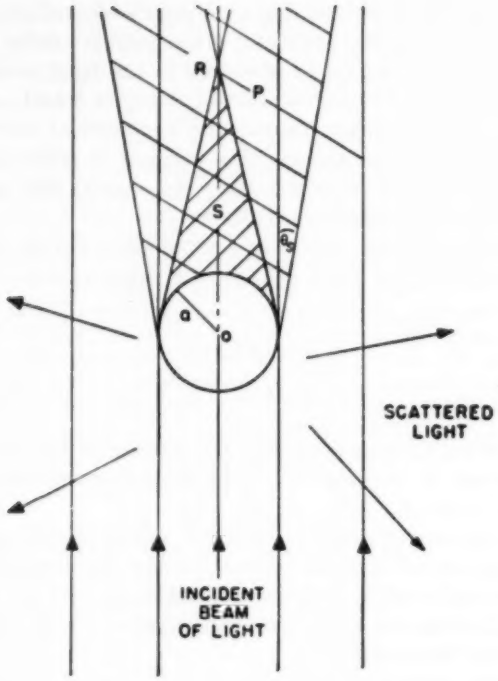


FIG. 2. Shadow structure for plane wave on a sphere.

Around the shadow, one should find a sort of penumbra P, exhibiting a succession of interference rings caused by the superposition of the incident field E_i and of the secondary field E_r. Outside the penumbra, the fields corresponding to direct wave and scattered wave could be distinguished from each other on account of their different polarization and direction of propagation.

When the wave-length λ is decreased indefinitely, the angle θ_s decreases and the penumbra region shrinks together, while the regular shadow extends farther and farther. More detailed explanations will be found at the end of the present paper and in Appendix III.

III. A RIGOROUS ELECTROMAGNETIC THEORY; THE PROBLEM OF THE FACTOR 2

The theory of light diffraction is usually built along the following lines. An incident wave (field E_i) falls upon an obstacle (a sphere) and the resulting field distribution is analyzed as the sum of two terms:

$$\mathbf{E}=\mathbf{E}_i+\mathbf{E}_r, \tag{6}$$

where $\mathbf{E}_i$ is the incident wave, as it would be obtained without perturbation if the sphere was not scattering any radiation, and $\mathbf{E}_r$ is the secondary wave, including the scattered radiation and the corrective field in the shadow region.

The energy flow for the scattered radiation is computed by taking the Poynting vector corresponding to E_r and integrating over all directions. The resulting energy flow is expressed by the scattering cross section Q_s, the area which receives equal energy flow from the incident waves. The theory is in the literature.[10] The computation of the electromagnetic field around the scattering sphere is given with all details by Stratton[2] (this part being essentially a summary of the original papers by Debye and Mie) and is absolutely correct. When the general formulas are used in connection with the assumption $\lambda<<a$ (λ is the wave-length, a the radius of the sphere), the scattering cross section is found to be twice the cross section of the sphere

$$Q_s\rightarrow 2\pi a^2, \quad \lambda<<a, \tag{7}$$

as will be shown in more detail in the following sections. This result was first explicitly stated in a paper by Stratton and Houghton.[11] Their Fig. 1 is a graph of the variation of a coefficient

$$K=\frac{|A|}{2\pi a^2},$$

where $|A|$ is the total energy scattered, which tends to the limit *one* for large values of a/λ. That is the result stated in (7).

Comparing formulas (2) and (7) we note a *discrepancy by a factor 2* in a problem where the sphere is very much larger than the wave-length.

This discrepancy is due to a real difficulty in the definitions. Let us *first* assume the validity of *geometrical optics*, with its results (2) and investigate their translation in the more elaborate method sketched in equation (6). The resultant field at any point outside the sphere is the sum of the primary and secondary field

$$\mathbf{E}=\mathbf{E}_i+\mathbf{E}_r,$$
$$\mathbf{H}=\mathbf{H}_i+\mathbf{H}_r$$

where $\mathbf{E}_i$, $\mathbf{H}_i$ are the electric and magnetic field as they would be found in the incident beam of light if it could propagate without perturbation. $\mathbf{E}_r$, $\mathbf{H}_r$ are the additional field components resulting from the perturbation due to the sphere. Stratton also calls this the "reflected field," hence the subscript r.

This is where a dubious physical picture is introduced. The secondary field $\mathbf{E}_r$, $\mathbf{H}_r$ contains two different things: the *reflected or scattered wave*, spreading all around the sphere in all directions (let us call it $\mathbf{E}_{r1}$, $\mathbf{H}_{r1}$) and the *shadow* ($\mathbf{E}_{r2}$, $\mathbf{H}_{r2}$). This second part is a field distribution which must compensate almost completely (but for some interference fringes) the unperturbed incident field $\mathbf{E}_i$, $\mathbf{H}_i$ in a region behind the sphere. Because $\mathbf{E}_i$, $\mathbf{H}_i$ is taken as the unperturbed incident wave, the secondary field $\mathbf{E}_{r2}$, $\mathbf{H}_{r2}$ must contain the correction to the incident wave in the region known as the shadow of the sphere.

$$\mathbf{E}_r=\mathbf{E}_{r1}+\mathbf{E}_{r2},$$
$$\mathbf{H}_r=\mathbf{H}_{r1}+\mathbf{H}_{r2}, \tag{8}$$

[10] See reference 2, p. 563.

[11] J. A. Stratton and H. G. Houghton, "A theoretical investigation of the transmission of light through fog," Phys. Rev. **38**, 159–165 (1931).

where the secondary field = scattered radiation + compensating field in the shadow.

The first point being clear, let us come back to the Stratton-Mie discussion. The radial component of the complex Poynting vector is[12]

$$S_R^* = \tfrac{1}{2}(E_\theta \bar{H}_\phi - E_\phi \bar{H}_\theta) \tag{9}$$

where the bar over the letter means complex conjugate. If we compute the field radiated at very large distance from the sphere, we may use the well-known result

$$H_\phi \approx \eta E_\theta, \quad H_\theta \approx -\eta E_\phi, \quad \eta = \left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}, \tag{10}$$

where η is the characteristic impedance of free space. Introducing the transverse component E_t of the electric field, we write

$$S_R^* = \frac{\eta}{2}[|E_\theta|^2 + |E_\phi|^2] = \frac{\eta}{2}|E_t|^2, \tag{11}$$

a relation which checks with Stratton.[9] This expression can be resolved by virtue of (6) into three groups

$$S_R^* = \frac{\eta}{2}[|E_{ti}|^2 + |E_{tr}|^2 + 2(E_{ti} \cdot \bar{E}_{tr})]. \tag{12}$$

The first term contains the field of the unperturbed incident wave, the second term represents the contribution of the secondary wave, and the last one the interaction between both fields. Let us now integrate over a sphere at large distance (radius $R >> a$). We notice immediately that the first term does not contribute anything,* while the integral of S_R^* yields the total energy absorbed by the sphere, hence the following relation[13]

$$W_a = -\mathrm{Re}\int\int S_R^* R^2 \sin\theta d\theta d\phi = -W_r + W_{ir},$$

where

$$W_r = \mathrm{Re}\frac{\eta}{2}\int\int |E_{tr}|^2 R^2 \sin\theta d\theta d\phi,$$

(the W_s of Stratton),

$$W_{ir} = -2\,\mathrm{Re}\frac{\eta}{2}\int\int (E_{ti} \cdot \bar{E}_{tr}) R^2 \sin\theta d\theta d\phi,$$

(the W_t of Stratton). (13)

Here the subscripts have been changed to r and ir, instead of s and t (Stratton) to show the difference with Stratton about the physical interpretation of the different terms.

The point of view of Stratton is the following: "The second term [in E_{tr}] *obviously* measures the outward flow of the secondary or scattered energy from the diffracting sphere, and the total scattered energy is W_s. W_t measures, therefore, the total energy derived from the primary wave and dissipated as heat and scattered radiation." As a matter of fact, for a metallic sphere $W_a = 0$ and $W_s = W_t$.

The explanations given in Eq. (8) clearly show that the Mie-Stratton interpretation of the different terms is not quite correct in the case of geometrical optics and that it rests on an incorrect physical picture. This will be investigated on a specific example to explain how it makes for the factor 2. Let us consider a metallic sphere of radius a much larger than the wave-length. The approximations of geometrical optics yield the following results.

A. *Scattered radiation.* The tangential component is

$$E_{tr1} = \frac{C}{R}E_{i0}, \quad |C^2| = \frac{a^2}{4},$$

where E_{i0} is the amplitude of the electric field in the incident wave. This results from the well-known fact that a reflecting sphere scatters light uniformly in all directions and that the total amount of energy scattered per second amounts to the total incident energy falling per second on the cross section πa^2 (geom optics only)

$$W_{\text{scattered}} = R^2\frac{\eta}{2}\int\int |E_{tr1}|^2 \sin\theta d\theta d\phi = \frac{\eta}{2}4\pi|C|^2 E_{i0}^2$$

$$= \pi a^2 \frac{\eta}{2}E_{i0}^2 = \pi a^2 S_2,$$

$$S_2 = \frac{\eta}{2}E_{ii}^2, \tag{14}$$

where S_2 is the mean energy flux in the incident wave.

B. *Shadow.* The field of the incident wave is completely compensated over an area πa^2 behind the sphere, hence a compensating field

$$E_{tr2} = -E_{i0} \text{ in the shadow } \pi a^2. \tag{15}$$

According to (8), the secondary field E_r is the sum of these two terms, and

$$|E_{tr}|^2 = |E_{tr1} + E_{tr2}|^2 = |E_{tr1}|^2 + |E_{tr2}|^2 + 2(E_{tr1} \cdot \bar{E}_{tr2}). \tag{16}$$

Integrating over a sphere at large distance we notice that the last term of (16) gives practically zero, since at large distance E_{tr1} decreases as $1/R$ while the area of integration $R^2 \sin\theta\Delta\theta\Delta\phi$ keeps a constant value πa^2

[12] See reference 2, p. 568, Eqs. (17) and (18).
[13] See reference 2, p. 568, Eqs. (20) and (21).

* Since as much incident energy falls on the sphere as leaves it

(area of the shadow). Hence

$$W_r=\frac{\eta}{2}\int\int|E_{tr_1}|^2R^2\sin\theta d\theta d\phi + \frac{\eta}{2}\int\int|E_{tr_2}|^2R^2\sin\theta d\theta d\phi$$

$$W_r=W_{\text{scattered}}+\pi a^2\frac{\eta}{2}|E_{tr_2}|^2=2\pi a^2\frac{\eta}{2}E_{i0}^2 \tag{17}$$

using (14) and (15). The flux of the negative field (15) over the area πa^2 of the shadow obviously gives an additional contribution just equal to that of the scattered radiation.

Let us now turn to the mixed term W_{ir} [Eq. (13)]. In this integral the scalar product $(E_{ti}\cdot\bar{E}_{tr})$ splits into

$$(E_{ti}\cdot\bar{E}_{tr})=(E_{ti}\cdot\bar{E}_{tr_1})+(E_{ti}\cdot\bar{E}_{tr_2}).$$

The first one is zero in average, and the second one is to be integrated over the area of the shadow

$$W_{ir}=-\eta\int(E_{ti}\cdot\bar{E}_{tr_2})dS=+\pi a^2\eta E_{i0}^2=W_r. \tag{18}$$

So we reach the following conclusion: In the case of a well-defined geometrical black shadow, the assumptions of Mie and Stratton yield a scattered energy twice as large as the actual scattering.

IV. QUALITATIVE DISCUSSION BASED ON PHYSICAL OPTICS

The case of geometrical optics, which we have just discussed, obviously represents a dangerous oversimplification of the real problem, and we immediately feel that there is no such thing as the sharply defined shadow of the geometrical theory.

Referring to Eqs. (3), (4), (5), and Fig. 2, we note that the actual shadow limit must extend from $\theta=0$ up to an angle θ_s, which may become small when $a>>\lambda$ but can never be exactly zero.

Geometrical optics leads to the conclusion that (1) energy scattered all around the sphere corresponds to a cross section πa^2 equal to the geometrical cross section; (2) an equal amount of secondary energy flow is found exactly on the angle $\theta=0$, thus yielding an additional πa^2 contribution to the cross section. The total makes $2\pi a^2$ and checks with Stratton's result in Eq. (7) above.

According to physical optics, we understand that the first contribution should remain unchanged and that the second one should spread over a small but finite angle θ_s.

The theoretical problem can now be stated in the following way. A plane electromagnetic wave with its electric vector in the x direction is propagating upwards along the z axis. It strikes a sphere centered on the origin O of coordinates and is scattered around the sphere. If the sphere is large enough to produce a shadow, the shadow will extend about the positive z axis. Using spherical coordinates (Fig. 1) R, θ, ϕ ($R=a$, radius of the sphere), the shadow corresponds to angles $O<\theta<\theta_s$, where θ_s is a small quantity which is to be conveniently chosen.

According to the preceding discussion, it is necessary to compute first the amount of energy scattered in all directions, except for the shadow region;

$$\begin{aligned} &R>>a \text{ far away from the sphere,}\\ &\theta_s<\theta\leqslant\pi \text{ outside the shadow,}\\ &O<\phi\leqslant 2\pi \text{ all around the sphere.}\end{aligned} \tag{19}$$

The problem is to take the expressions for the Poynting vector, as obtained by Stratton[2] and to carry the integration of energy flow according to conditions (19), instead of doing it for all angles $0<\theta\leqslant\pi$ as Stratton did. This last procedure is objectionable since it includes the shadow region, and incorporates into the scattered field radiation the whole field distribution inside the shadow, which is physically to be distinguished from the scattered radiation.

V. EXPERIMENTAL CHECKS OF THE THEORY

As stated in the first section, the present investigation was prompted by some discrepancies between experiment and theory. Those who discovered such discrepancies also found another source of error, caused by inconsistent definitions for the intensity of a beam of light. The errors caused by inconsistent use of different systems of units are pointed out in AMP Report 87.2R.[14]

When the preceding errors have been corrected, the Mie-Stratton theory is found to work very well for all cases

$$\lambda\gtrsim\tfrac{1}{2}a,$$

$$\rho_1=2\pi\frac{a}{\lambda}<15, \tag{20}$$

and experimental results check very well with theoretical formulas. The values of the total scattering, observed in LaMer's laboratory for sulfur or oleic acid droplets, agree very well with the theory.

For very small spheres and long wave-length ($\lambda>>a$) the theory still gives the correct result for Rayleigh scattering, as pointed out by Stratton.[15]

The problem discussed in Sections II and III is related to the case of large ρ_1 values

$$\rho_1>15,\quad \lambda<\tfrac{1}{2}a. \tag{21}$$

In the following experiments by Sinclair, for instance,

$$\begin{aligned} a&\approx 15\mu,\\ \lambda&\approx 0.5\mu,\\ \rho_1&\approx 190.\end{aligned} \tag{22}$$

[14] Leon Brillouin, "On light scattering by spheres. II," AMP Report 87.2R, (OEMsr-1007), OSRD 3464, AMG-C, April 1944, AMP-203-M1.

[15] See reference 2, p. 572.

There, according to experimental conditions, the observed scattering cross section is πa^2 when the limit angle θ_s (see Section III) is larger than 10 degrees and goes up to $2\pi a^2$ for smaller θ_s values as obtained for large distances R of observation.

The following account, written by Dr. Sinclair, summarizes the experimental checks:

> We have recently made a few observations of the size of lycopodium spores of fairly uniform size. The radius of these spores, as measured in a light microscope, was found to be 15.0±0.9 micron.
>
> The spores were allowed to settle on a glass plate which was then placed in a beam of light of parallel rays and in a slightly divergent beam from a point source. Measurements were made of the angular diameter of two orders of both the light and dark diffraction rings in both red and green light. From these measurements we calculated the spore radius to be 15.5±0.2 micron, in good agreement with the microscope measurements.
>
> We also placed the plate of lycopodium spores in a parallel beam in front of a photocell. The reading of the photocell increased by the factor 2 to 1 when the distance from the plate to the photocell was decreased from 18 feet to 6 inches. This result was predicted by Professor Brillouin. Calculation checked this result and showed that at 18 feet the observed scattering cross section was $2\pi a^2$ and at 6 inches $a^2\pi$. In other words, the true scattering cross section is $2\pi a^2$.

VI. STRATTON'S PRESENTATION OF THE MIE THEORY

The general method[16] is to expand the incident plane wave into an infinite series of terms, each of which corresponds to a specific spherical harmonic, then to assume a similar expansion with arbitrary coefficients for the scattered or secondary wave and to obtain the coefficients by matching both fields on the surface of the sphere to satisfy boundary conditions.

The solution is thus obtained as an expansion in spherical harmonics and would be perfectly suitable if we were in a position to observe separately each term in the series. Unfortunately this is not the case, and we are only able to measure the total field scattered in any given direction. Hence we are faced with the very difficult task of rebuilding a function that is given as a rather clumsy expansion.

The point of departure is the set of formulas given by Stratton.[17] The incident wave is taken as

$$E_{ix}=E_0e^{-i\omega t+ikz},$$

$$H_{iy}=-\eta_0E_0e^{-i\omega t+ikz}, \tag{23}$$

$$k=\frac{2\pi}{\lambda}=\frac{\omega}{c}=\omega(\epsilon_0\mu_0)^{\frac{1}{2}}, \quad \eta_0=(\epsilon_0/\mu_0)^{\frac{1}{2}}, \tag{24}$$

where ϵ_0 and μ_0 are the permittivity and permeability in the surrounding medium (free space), in m.k.s. units. The sphere has its center at O and a radius a. Its electric and magnetic properties are specified by the values ϵ_1, μ_1, which may be complex if necessary to include absorption, and

$$k_1=\omega(\epsilon_1\mu_1)^{\frac{1}{2}}. \tag{25}$$

The solution of the problem is obtained by writing the continuity conditions on the surface of the sphere. When this is done, the field outside the sphere comes out as the sum of two terms,

$$\mathbf{E}=\mathbf{E}_i+\mathbf{E}_r, \quad \mathbf{H}=\mathbf{H}_i+\mathbf{H}_r, \tag{26}$$

where $\mathbf{E}_i$, $\mathbf{H}_i$ are the fields corresponding to the unperturbed incident wave (23) while $\mathbf{E}_r$, $\mathbf{H}_r$ represent the corrective terms, which contain the reflected or scattered waves spreading out from the sphere and the corrective field which compensates the incident wave in the shadow region.

Stratton gives[18] the complete formulas for the secondary fields $\mathbf{E}_r$, $\mathbf{H}_r$, which appear as a sum of terms corresponding to the radiation of electric n poles (amplitude b_n) or magnetic n poles (amplitude a_n). The field components of the secondary field are

$$E_{rR}=E_0e^{-i\omega t}\sum_{n=1}^{\infty}i^{n-1}\frac{(2n+1)}{\rho}b_nh_n(\rho)P_n^1\cos\phi,$$

$$E_{r\theta}=E_0e^{-i\omega t}\sum_{n=1}^{\infty}i^n\frac{2n+1}{n(n+1)}\left[a_n\frac{h_n}{\sin\theta}P_n^1-ib_n\frac{1}{\rho}(\rho h_n)'\frac{\partial P_n^1}{\partial\theta}\right]\cos\phi,$$

$$E_{r\phi}=E_0e^{-i\omega t}\sum_n i^n\frac{2n+1}{n(n+1)}\left[-a_nh_n\frac{\partial P_n^1}{\partial\theta}+ib_n\frac{1}{\rho}(\rho h_n)'\frac{P_n^1}{\sin\theta}\right]\sin\phi,$$

$$H_{rR}=-\eta_0E_0e^{-i\omega t}\sum_n i^{n+1}\frac{2n+1}{\rho}a_nh_nP_n^1\sin\phi,$$

$$H_{r\theta}=-\eta_0E_0e^{-i\omega t}\sum_n i^n\frac{2n+1}{n(n+1)}\left[-b_n\frac{h_n}{\sin\theta}P_n^1+ia_n\frac{1}{\rho}(\rho h_n)'\frac{\partial P_n^1}{\partial\theta}\right]\sin\phi,$$

$$H_{r\phi}=-\eta_0E_0e^{-i\omega t}\sum_n i^n\frac{2n+1}{n(n+1)}\left[-b_nh_n\frac{\partial P_n^1}{\partial\theta}+ia_n\frac{1}{\rho}(\rho h_n)'\frac{\dot{P}_n^1}{\sin\theta}\right]\cos\phi, \tag{27}$$

where the prime indicates a ρ-derivative and $\rho=kR=(2\pi R)/\lambda$.

It may be noted here that the same formulas can be used for the representation of the incident plane wave

[16] See reference 2, p. 409.
[17] See reference 2, pp. 564–572.
[18] See reference 2, p. 564.

as a superposition of waves from electric and magnetic multipoles:

> Plane wave polarized along x and propagating along z: take $a_n=b_n=1$ and replace $h_n(\rho)$ by $j_n(\rho)$. (27a)

The only trouble about these series is their very poor convergence. If the equations in series (27) are used according to (27a) to represent a plane wave, the terms of importance run up to $n>\rho=kR$, which is a very large figure.

Coming back to the scattered radiation (27), we shall find a simplification, since the coefficients a_n and b_n become very small when $n>\rho_1=ka$ and may often be neglected. In order to compute the scattered energy, all we need is the field at a great distance R,

$$R>>a, \quad \rho>>\rho_1, \quad \rho=kR, \quad \rho_1=ka, \quad k=2\pi/\lambda, \quad (28)$$

hence

$$kR>>n.$$

This condition enables us to use Debye's approximate formulas[19] and to take the asymptotic expressions of the spherical Bessel functions

$$h_n^1(\rho)=\left(\frac{\pi}{2\rho}\right)^{\frac{1}{2}}H_{n+\frac{1}{2}}(\rho)\approx\frac{1}{\rho}e^{i(\rho-[(n+1)/2]\pi)},$$

$$j_n(\rho)=\left(\frac{\pi}{2\rho}\right)^{\frac{1}{2}}J_{n+\frac{1}{2}}(\rho)\approx\frac{1}{\rho}\cos\left(\rho-\frac{n+1}{2}\pi\right)$$

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}[\rho h_n^1]\approx\frac{i}{\rho}e^{i(\rho-[(n+1)/2]\pi)}=ih_n^1. \quad (28a)$$

If these simplified expressions are used in Stratton's fundamental equations (27), the secondary field at a great distance is obtained. The factor i^n in equation (27) partly compensates $e^{-i[(n+1)/2]\pi}$ from (28a),

$$i^n e^{-i[(n+1)/2]\pi}=-i,$$

and we find

$$E_{rR}=H_{rR}=0,$$

$$E_{r\theta}=\frac{1}{\eta_0}=-H_{r\phi}=\frac{-i}{\rho}E_0e^{-i\omega t+i\rho}\sum_n\frac{2n+1}{n(n+1)}\times\left(a_n\frac{P_n^1}{\sin\theta}+b_n\frac{dP_n^1}{d\theta}\right)\cos\phi,$$

$$E_{r\phi}=\frac{-1}{\eta_0}=H_{r\theta}=\frac{i}{\rho}E_0e^{-i\omega t+i\rho}\sum_n\frac{2n+1}{n(n+1)}\times\left(a_n\frac{dP_n^1}{d\theta}+b_n\frac{P_n^1}{\sin\theta}\right)\sin\phi. \quad (29)$$

[19] See reference 2, pp. 568–569, Eq. (22), p. 404, Eq. (23).

The field at a great distance is transverse and satisfies the usual relation

$$\mathbf{H}=\eta_0(\mathbf{I}_1\times\mathbf{E}), \quad (30)$$

where $\mathbf{I}_1$ is a unit vector in the R direction of propagation. The complex Poynting's vector is directed along R

$$S_R^*=\tfrac{1}{2}(E_{r\theta}\bar{H}_{r\phi}-E_{r\phi}\bar{H}_{r\theta})=\tfrac{1}{2}\eta_0(|E_{r\theta}|^2+|E_{r\phi}|^2), \quad (31)$$

where the bar over the letter means "complex conjugate." As stated before, the secondary field [Eq. (29)] represents the scattered radiation plus the compensating field in the shadow. What we want to compute is the energy radiated in all directions, shadow excluded. This can be done by integrating the Poynting's vector for angles θ ranging from a small but finite value θ_s up to π.

$$W_s=\mathrm{Re}\int_0^{2\pi}d\phi\int_{\theta_s}^{\pi}d\theta S_R^*R^2\sin\theta, \quad (32)$$

where Re means "Real part of . . . ." The amount of energy scattered is better expressed by giving the *scattering cross section* Q_s,

$$Q_s\tfrac{1}{2}\eta_0E_0^2=W_s, \quad (33)$$

which means that the amount W_s scattered is equal to the energy of the incident wave falling on an area Q_s. Finally

$$Q_s=\frac{\mathrm{Re}}{k^2}\int_0^{2\pi}d\phi\int_{\theta_s}^{\pi}d\theta\sum_n\sum_m\frac{2n+1}{n(n+1)}\frac{2m+1}{m(m+1)}\times\left[\left(a_n\frac{P_n^1}{\sin\theta}+b_n\frac{dP_n^1}{d\theta}\right)\left(\bar{a}_m\frac{P_m^1}{\sin\theta}+\bar{b}_m\frac{dP_m^1}{d\theta}\right)\cos^2\phi+\left(a_n\frac{dP_n^1}{d\theta}+b_n\frac{P_n^1}{\sin\theta}\right)\left(\bar{a}_m\frac{dP_m^1}{d\theta}+\bar{b}_m\frac{P_m^1}{\sin\theta}\right)\times\sin^2\phi\right]\sin\theta. \quad (34)$$

The ϕ integration is obvious,

$$\int_0^{2\pi}\cos^2\phi d\phi=\int_0^{2\pi}\sin^2\phi d\phi=\pi.$$

Hence

$$Q_s=\frac{\pi}{k^2}\mathrm{Re}\int_{\theta_s}^{\pi}d\theta\sum_n\sum_m\frac{2n+1}{n(n+1)}\frac{2m+1}{m(m+1)}\times\left\{(a_n\bar{b}_m+a_m\bar{b}_n)\left[P_n^1\frac{dP_m^1}{d\theta}+P_m^1\frac{dP_n^1}{d\theta}\right]+(a_n\bar{a}_m+b_n\bar{b}_m)\left[\frac{P_n^1P_m^1}{\sin\theta}+\sin\theta\frac{dP_n^1}{d\theta}\frac{dP_m^1}{d\theta}\right]\right\}. \quad (34a)$$

This is easily compared with Stratton's formula.[20]

[20] See reference 2, p. 569, Eq. (28).

Stratton takes $\theta_s=0$ in which case the first θ integral is zero and the second is different from zero only when $m=n$, where it yields

$$\int_0^\pi \left(\frac{P_n^1 P_n^1}{\sin\theta}+\sin\theta\frac{dP_n^1}{d\theta}\frac{dP_n^1}{d\theta}\right)d\theta=2\frac{n^2(n+1)^2}{2n+1}. \quad (35)$$

Hence

$$Q_{\text{Stratton}}=\frac{2\pi}{k^2}\sum_n (2n+1)(|a_n|^2+|b_n|^2). \quad (36)$$

Thus, taking an angle $\theta_s=0$ simplifies the formulas very much. But we need to keep a small angle $\theta_s\neq 0$ and the general expression (34a):

$$Q_s=\frac{\pi}{k^2}\sum_{n,\,m}\frac{2n+1}{n(n+1)}\frac{2m+1}{m(m+1)}\operatorname{Re}[I_1(a_n\bar{b}_m+a_m\bar{b}_n) +I_2(a_n\bar{a}_m+b_n\bar{b}_m)], \quad (37)$$

with

$$I_1=\int_{\theta_s}^\pi\left(P_n^1\frac{dP_m^1}{d\theta}+P_m^1\frac{dP_n^1}{d\theta}\right)d\theta=\left|P_n^1P_m^1\right|_{\theta_s}^\pi,$$

$$I_2=\int_{\theta_s}^\pi\left(\frac{P_n^1P_m^1}{\sin\theta}+\sin\theta\frac{dP_n^1}{d\theta}\frac{dP_m^1}{d\theta}\right)d\theta$$

$$=\left|\sin\theta P_m^1\frac{dP_n^1}{d\theta}\right|_{\theta_s}^\pi+n(n+1)\int_{\theta_s}^\pi P_n^1P_m^1\sin\theta d\theta. \quad (38)$$

[See Appendix I, Eq. (87).] These are our fundamental formulas for the following discussion.

VII. USE OF APPROXIMATE EXPRESSIONS FOR THE LEGENDRE FUNCTION

The problem is to obtain approximate values for the integrals I_1 and I_2 of (38). Using formula (93) of Appendix I for the lower limit θ_s and remembering P_n^1 to be zero for π, we obtain

$$I_1=P_n^1P_m^1\Big|_{\theta_s}^\pi\approx-(nm)^{\frac{1}{2}}\frac{2}{\pi\sin\theta_s}\sin\left[\left(n+\frac{1}{2}\right)\theta_s+\frac{3\pi}{4}\right]\sin\left[\left(m+\frac{1}{2}\right)\theta_s+\frac{3\pi}{4}\right],$$

$$I_1=\frac{(nm)^{\frac{1}{2}}}{\pi\sin\theta_s}[\sin(n+m+1)\theta_s-\cos(n-m)\theta_s] \quad (39)$$

when

$$n, m>>1,$$

where θ_s is small and $\sin\theta_s$ can be replaced by θ_s, while $(n+m+1)\theta_s$ and $(n-m)\theta_s$ may become rather large.

As for the integral I_2, we can apply the same treatment to the first term, but if we use approximate expressions for $P_n^1P_m^1$ in the second integral, we cannot extend it up to π but only to $\pi-\epsilon_2$, hence

$$I_2\approx\sin\theta P_m^1\frac{dP_n^1}{d\theta}\Big|_{\theta_s}^\pi+n(n+1)\int_{\theta_s}^{\pi-\epsilon_2}P_n^1P_m^1\sin\theta d\theta,$$

$$\approx-\frac{2n}{\pi}(nm)^{\frac{1}{2}}\sin\left[\left(m+\frac{1}{2}\right)\theta_s+\frac{3\pi}{4}\right]\cos\left[\left(n+\frac{1}{2}\right)\theta_s+\frac{3\pi}{4}\right]-\frac{(nm)^{\frac{1}{2}}}{\pi}n(n+1)\int_{\theta_s}^{\pi-\epsilon_2}[\sin(n+m+1)\theta-\cos(n-m)\theta]d\theta,$$

$$I_2\approx\frac{+n(nm)^{\frac{1}{2}}}{\pi}\Big[+\cos(n+m+1)\theta_s+\sin(n-m)\theta_s+(n+1)\left|\frac{\cos(n+m+1)\theta}{n+m+1}+\frac{\sin(n-m)\theta}{n-m}\right|_{\theta_s}^{\pi-\epsilon_2}\Big]. \quad (40)$$

The last term is the most important one, especially when n and m do not differ too much, and $n+1/n-m$ is very large. For $n=m$ we have

$$\left|\frac{\sin(n-m)\theta}{n-m}\right|_{\theta_s}^{\pi-\epsilon_2}=\left|\theta\right|_{\theta_s}^{\pi-\epsilon_2}=\pi-\theta_s-\epsilon_2, \qquad n=m>>1$$

$$I_2=\frac{n^2}{\pi}\Big[\cos(2n+1)\theta_s+\frac{1}{2}\Big|\cos(2n+1)\theta\Big|_{\theta_s}^{\pi-\epsilon_2} +n\pi-n\theta_s-n\epsilon_2\Big]\approx n^3. \quad (41)$$

According to our assumptions, n is supposed to be large compared to 1 and θ_s, ϵ_2 are small, but $n\theta_s$, $n\epsilon_2$ may be of the order of some units. Anyhow, in formula (41), the term $n\pi$ is by far the largest one in the brackets, and our approximate result compares favorably with the correct one (Stratton, $\theta_s=\epsilon_2=0$)

$$I_2=\frac{2n^2(n+1)^2}{2n+1}\approx n^3 \quad n>>1. \quad (42)$$

In our general formula (37), we note that the large values of n, m give the largest contribution, since integral I_2 is of the order of n^3 and even after multiplication by $\frac{2n+1}{n(n+1)}\frac{2m+1}{m(m+1)}$ it still increases with n. This remark enables us to write separately the term $n=m$, which practically coincides with the one kept by Stratton, and to state

$$Q_s=Q_{\text{Stratton}}+Q_{\text{corr}}$$

$$Q_{\text{corr}}=\frac{\pi}{k^2}\sum_n\sum_{n\neq m}\frac{2n+1}{n(n+1)}\frac{2m+1}{m(m+1)}\operatorname{Re}[2I_1a_n\bar{b}_m +I_2(a_n\bar{a}_m+b_n\bar{b}_m)]. \quad (43)$$

In the double sum on n and m we have replaced $a_n b_m + a_m b_n$ by its equivalent $2a_n b_m$. The quantity Q_{corr} yields the correction to Stratton's original formula for the scattering cross section.

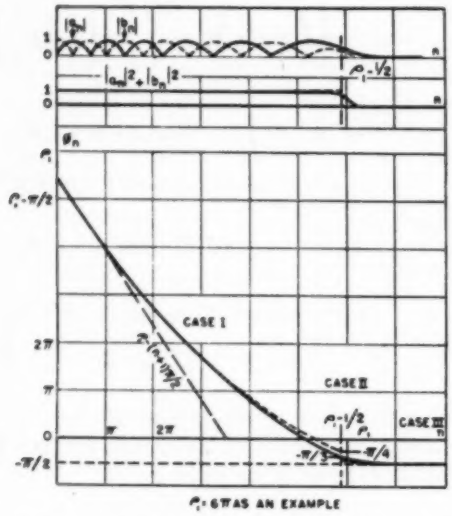


FIG. 3. Behavior of the phase angle and the coefficients.

VIII. LARGE METALLIC SPHERES; COMPUTATION OF THE COEFFICIENTS a_n, b_n

The preceding formulas should be used as point of departure for the discussion of the scattering cross section for large spheres.

$$\rho_1 = \frac{2\pi a}{\lambda} >> 1 \quad a \gtrsim \lambda. \tag{44}$$

The approximations involved will mean neglecting terms in $1/\rho_1$ compared with unity.

For a perfect metallic sphere, the a_n and b_n coefficients have the following values[21]

$$a_n = -\frac{j_n(\rho_1)}{h_n^1(\rho_1)}; \quad b_n = -\frac{[\rho_1 j_n]'}{[\rho_1 h_n^1]'}; \quad \rho_1 = ka = 2\pi\frac{a}{\lambda}, \tag{45}$$

where j_n, h_n^1 are the spherical Bessel functions, as in Eq. (28a) and the prime denotes the derivative with respect to ρ_1.

These formulas (45) may be written in a slightly different way, if we start from an expression

$$\rho_1 h_n^1 = A_n e^{i\phi_n}; \quad A_n \text{ real, positive,} \tag{46}$$

where amplitude A_n and phase angle ϕ_n of the h_n^1 functions are explicitly used. Then

$$\rho_1 j_n = \mathrm{Re}\rho_1 h_n^1 = A_n \cos\phi_n \tag{46a}$$

and

$$a_n = -\cos\phi_n e^{-i\phi_n}. \tag{47}$$

[21] See reference 2, p. 565, Eq. (13).

Taking the derivatives, we find

$$[\rho_1 h_n^1]' = [A_n' + iA_n\phi_n']e^{i\phi_n},$$
$$[\rho_1 j_n]' = \mathrm{Re}[\rho_1 h_n^1]' = A_n' \cos\phi_n - A_n\phi_n' \sin\phi_n.$$

Here, two cases should be considered separately

$$\text{(A)}\ \phi_n' >> \frac{A_n'}{A_n} \quad b_n \approx -i\sin\phi_n e^{-i\phi_n},$$

$$\text{(B)}\ \phi_n' << \frac{A_n'}{A_n} \quad b_n \approx -\cos\phi_n e^{-i\phi_n} = a_n. \tag{48}$$

These results are identical with those of Stratton,[22] except for the fact that Stratton uses an angle

$$\gamma_n = \frac{\pi}{2} - \phi_n.$$

The important point is that the a_n, b_n coefficients are large (order of magnitude 1) when $n \lesssim \rho_1$ and decrease very rapidly when $n > \rho_1$. In order to prove this and to find approximate expressions to be used in each case, we must distinguish three different cases:

I $\quad n << \rho_1 \quad \rho_1 >> 1$
II $\quad n \approx \rho_1$
III $\quad n > \rho_1$

and discuss the corresponding approximate expressions given by Debye.[23]

Case I. For $(n << \rho_1 >> 1)$, the approximate formulas to be used are

$$n << \rho_1, \tag{49}$$

$$A_n \approx 1, \quad \phi_n \approx \rho_1 - \frac{\pi}{2}(n+1); \tag{50}$$

$$\rho_1 j_n \approx \cos\phi_n, \quad \rho_1 h_n^1 \approx e^{i\phi_n}. \tag{51}$$

In Eq. (46) the amplitude A_n varies very slowly and we may use formulas (46) and (47),

$$a_n = -\cos\phi_n e^{-i\phi_n}$$
$$b_n = -i\sin\phi_n e^{-i\phi_n}, \tag{52}$$

from which we compute

$$a_n \bar{b}_m = -i\cos\phi_n \sin\phi_m e^{i(\phi_m - \phi_n)},$$
$$a_n \bar{a}_m + b_n \bar{b}_m = \cos(\phi_m - \phi_n) e^{i(\phi_m - \phi_n)}. \tag{53}$$

Case II. For $n \approx \rho_1 >> 1$, the approximate expressions for the Bessel spherical functions are more complicated and have been computed.[23,24] They yield

$$\rho_1 = n + \frac{1}{2}, \quad \phi_n = -\frac{\pi}{3}. \tag{54}$$

[22] See reference 2, p. 565, Eq. (14).
[23] P. Debye, "Der Lichtdruck auf Kugeln von Beliebigem Material," Ann. d. Physik **30**, 57–136 (1909).
[24] G. Jobst, "Zur Farben theorie Kolloidaler Metallsuspensionen," Ann. d. Physik **76**, 863–888 (1925).

Case III. For $n>\rho_1>>1$, the approximate expressions given by Debye must be corrected to include the real part ψ_n in ζ_n, which Debye drops out. It is actually true that ψ_n is much smaller than the imaginary part of ζ_n, but it is safer to keep it, and to write

$$\zeta_n=\rho_1 h_n^1=\psi_n-i\frac{e^{\rho_1 f_n^1}}{(\sinh\tau_n^1)^{\frac{1}{2}}}, \tag{55}$$

$$\psi_n=\rho_1 j_n=\frac{1}{2}\frac{e^{-\rho_1 f_n^1}}{(\sinh\tau_n^1)^{\frac{1}{2}}}<<\zeta_n, \tag{56}$$

where

$$f_n^1=-\sinh\tau_n^1+\tau_n^1\cosh\tau_n^1>0, \quad \tau_n^1>0 \text{ real}, \tag{57}$$

$$\cosh\tau_n^1=\frac{n+\frac{1}{2}}{\rho_1}. \tag{58}$$

The factor $\frac{1}{2}$ in ψ_n (or j_n) was first omitted by Debye and discovered later by F. Emde.[25] The phase angle ϕ_n in ζ_n is given by

$$\tan\phi_n=-2e^{2\rho_1 f_n^1}, \tag{59}$$

which, for the lower limit of the interval $(n=\rho_1-\frac{1}{2})$, yields $\tan\phi_n=-2$, or ϕ_n slightly below $-\pi/3$, while for increasing n the angle ϕ_n reaches $-\pi/2$ as

$$\phi_n=-\frac{\pi}{2}+\frac{1}{2}e^{-2\rho_1 f_n^1}. \tag{60}$$

Under such conditions, we may use formulas (47) and (48) to compute the coefficients a_n and b_n,

$$a_n\approx b_n\approx-\cos\phi_n e^{-i\phi_n}\approx\frac{i}{2}e^{-2\rho_1 f_n^1}. \tag{61}$$

When n increases, both a_n and b_n decrease very rapidly and tend exponentially to zero,

$$n>>\rho_1,$$

$$\cosh\tau_n^1\approx\sinh\tau_n^1\approx\tfrac{1}{2}e^{+\tau_n^1}=\frac{n+\frac{1}{2}}{\rho_1}, \tag{62}$$

$$\rho_1 f_n^1=(n+\tfrac{1}{2})(\tau_n^1-1)=(n+\tfrac{1}{2})\left(\log\frac{2n+1}{\rho_1}-1\right)$$

$$=(n+\tfrac{1}{2})\log\frac{2n+1}{e\rho_1}, \tag{63}$$

$$\zeta_n=-i\left(\frac{2}{e}\right)^{\frac{1}{2}}\left(\frac{2n+1}{e\rho_1}\right)^n\to\infty,$$

$$\psi_n=\frac{1}{(2e)^{\frac{1}{2}}}\left(\frac{e\rho_1}{2n+1}\right)^{n+1}\to 0, \tag{64}$$

$$a_n=b_n=\frac{i}{2}\left(\frac{e\rho_1}{2n+1}\right)^{2n+1}. \tag{65}$$

Following the same line as in Case I, we compute the values of the combinations we need in our formulas, namely

$$a_n\bar{b}_m=\tfrac{1}{4}e^{-2\rho_1(f_n^1+f_m^1)}\to\frac{1}{4}\left(\frac{e\rho_1}{2n+1}\right)^{2n+1}\left(\frac{e\rho_1}{2m+1}\right)^{2m+1},$$

$$a_n\bar{a}_m+b_n\bar{b}_m=\tfrac{1}{2}e^{-2\rho_1(f_n^1+f_m^1)}\to\frac{1}{2}\left(\frac{e\rho_1}{2n+1}\right)^{2n+1}\times\left(\frac{e\rho_1}{2m+1}\right)^{2m+1}. \tag{66}$$

It is interesting to draw a few curves in order to visualize the results just obtained. As noticed in (47) and (48), we are especially interested in the phase angle ϕ_n, from which our a_n, b_n coefficients can be readily computed.

In Case I $(n<\rho_1-\frac{1}{2})$ the phase angle ϕ_n decreases regularly from $\rho_1-(\pi/2)(n=0)$ down to $-\pi/4(n=\rho_1-\frac{1}{2})$, this last point being reached with a horizontal tangent, but we have to stop using the formulas of Case I before we reach this limit, which corresponds to Case II. There the angle ϕ_n exhibits a slow decrease with $-\pi/3$ as its actual value when $n=\rho_1-\frac{1}{2}$. For larger values of n, we turn to the formulas of Case III, which show a decrease from $-\pi/3$ to $-\pi/2$. These results are summarized on the curve of Fig. 3, which should be considered only as qualitative.

From the ϕ_n curve, one may readily deduce the variation of a_n, b_n. The curves on top of Fig. 3 show the behavior of $|a_n|$ and $|b_n|$, which oscillate like $|\cos\phi_n|$ and $|\sin\phi_n|$ in Region I, then decrease in Region II and tend exponentially to zero in Region III.

The sum $|a_n|^2+|b_n|^2$ is equal to 1 in Region I, decreases rapidly in II and is practically zero in III

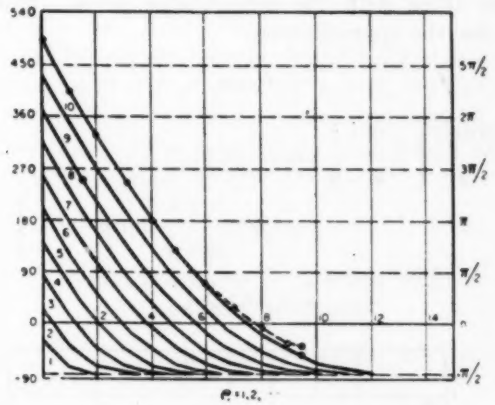


FIG. 4. The phase angle for various values of ρ_1.

(second curve on top of Fig. 3). This typical behavior had been noticed by Debye, who consequently replaced the curve by a rectangular one:

$$|a_n|^2+|b_n|^2=\begin{cases}1 & n<\rho_1-\frac{1}{2}\\ 0 & n>\rho_1-\frac{1}{2}.\end{cases}$$

When this approximation is used in Eq. (36), it yields

$$Q_{\text{Stratton}} \approx \frac{2\pi}{k^2} \sum_{n=1}^{n=\rho_1} (2n+1) = \frac{2\pi}{k^2} ka(ka+2) \approx 2\pi a^2,$$

since $\rho_1 = ka$. This shows directly how Stratton's scattering cross section comes out twice the actual cross section [Eq. (7)].

Some quantitative data are summarized by the curves of Fig. 4, which have been plotted from the Lowan tables "Scattering angles for spherical coordinates."

The circles, just near the curve $\rho_1 = 10$, correspond to points computed from Debye's formulas. They lie very near the curve, a result which suggests that for $\rho_1 > 10$ these asymptotic expressions can already be used for practical computation.

IX. COMPUTATION OF THE SCATTERING CROSS SECTION

The formulas obtained in the preceding sections must now be used for a discussion of the scattering cross section of a metallic sphere. We start from formula (43),

$$Q_s = Q_{\text{Stratton}} + Q_{\text{corr}}, \tag{67}$$

where Q_{Stratton} is the value obtained by Stratton, which tends to the limit $2\pi a^2$ for large spheres, and Q_{corr} is the correction, which should amount to $-\pi a^2$ for large spheres, in order to bring Q_s back to the expected value πa^2. Formula (43) gives Q_{corr} and we take I_1 and I_2 from Eqs. (39) and (40). Remembering that n and m may practically take all values up to ρ_1, while θ_s is small but finite, we note that I_1 is certainly much smaller than I_2 and that the most important terms in I_2 are those with the denominator $n-m$. Hence we may use the approximation

$$I_1 \approx 0, \tag{68}$$

$$I_2 \approx \frac{n(nm)^{\frac{1}{2}}}{\pi}\left[-\frac{m+1}{m-n}\sin(m-n)\theta_s - (-1)^{m-n}\frac{n+1}{m-n}\sin(m-n)\epsilon_2\right], \tag{69}$$

$$\rho_1 = ka >> 1.$$

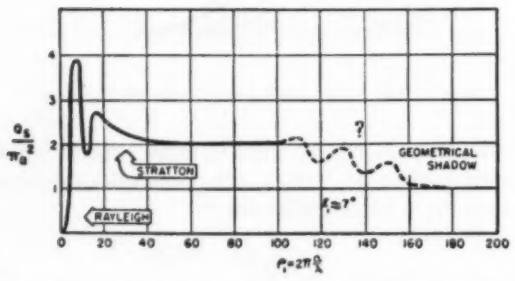


FIG. 5. Qualitative representation of scattering cross section.

The difference in the behavior of the terms in θ_s and ϵ_2 is striking. Terms in θ_s may pile up to a finite value, since they all keep the same sign, while alternative signs on the ϵ_2 terms will prevent them from yielding any large contribution (see Appendix II). We shall be satisfied with keeping the θ_s terms only, and we may introduce a new integer

$$\nu = m - n \neq 0 \quad \nu = +1, +2, +3 \cdots \tag{70}$$

$$Q_{\text{corr}} \approx -\frac{1}{k^2}\sum_n \frac{(2n+1)^2}{n+1} \sum_{\nu \neq 0} \left(1+\frac{2\nu}{2n+1}\right)\left(1+\frac{\nu}{n}\right)^{-\frac{1}{2}} \times \frac{\sin\nu\theta_s}{\nu} \operatorname{Re}(a_n \bar{a}_{n+\nu} + b_n \bar{b}_{n+\nu}). \tag{71}$$

As a further approximation, we assume that the most important contribution will be given by small ν values, and we put

$$|\nu| << n, \quad 1+\frac{2\nu}{2n+1} \approx 1+\frac{\nu}{n} \approx 1.$$

Then we use formula (53), which refers to the most important case, $n < \rho_1$, as discussed in Section VIII, and we simply drop all terms for which n or $n+\nu$ is larger than ρ_1

$$\operatorname{Re}(a_n \bar{a}_{n+\nu} + b_n \bar{b}_{n+\nu}) = \cos^2(\phi_{n+\nu} - \phi_n) \tag{72}$$

$$Q_{\text{corr}} \approx -\frac{1}{k^2}\sum_{n=1}^{\rho_1} \frac{(2n+1)^2}{n+1} \sum_{\nu \neq 0} \frac{\sin\nu\theta_s}{\nu} \cos^2(\phi_{n+\nu} - \phi_n). \tag{73}$$

The $\cos^2(\phi_{n+\nu} - \phi_n)$ is $\frac{1}{2}$ on the average* and

$$\sum_{\nu>0} \frac{\sin\nu\theta_s}{\nu} = \frac{\pi - \theta_s}{2} \quad \theta_s > 0. \tag{74}$$

[See Appendix II, Eq. (95).] The sum for negative ν values yields the same result, hence

$$\sum_{\nu \neq 0} \frac{\sin\nu\theta_s}{\nu} = \sum_{\nu>0} + \sum_{\nu<0} = \pi - \theta_s, \tag{75}$$

and

$$Q_{\text{corr}} = -\frac{\pi - \theta_s}{2k^2} \sum_{n=1}^{n=\rho_1} \frac{(2n+1)^2}{n+1} \approx -2\frac{\pi - \theta_s}{k^2} \sum_{n=1}^{\rho_1} n \approx -2\frac{\pi - \theta_s}{k^2}\frac{\rho_1^2}{2}. \tag{76}$$

We may assume θ_s small compared to π and obtain finally

$$Q_{\text{corr}} \approx -\pi\frac{\rho_1^2}{k^2} = -\pi a^2, \tag{77}$$

which is the expected result.

* For a more precise discussion, see end of Appendix II. $\phi_{n+\nu} - \phi_n$ for small ν values is practically $\nu(\partial\phi/\partial n)$, and $\partial\phi/\partial n$ is comprised between $-\pi/2$ and $-\pi/6$, as shown on Figs. 3 and 4.

A discussion of the approximations involved in this last step would now appear desirable.

The computation rests on the approximate formula for the Legendre function [Appendix I, Eq. (93)], as used in Section VII, and this approximation is good only when the small angle θ_s is larger than $4/n$ [Appendix I, Eq. (94)].

$$\theta_s \geqslant \frac{4}{n} \qquad n>>1. \quad (78)$$

On the other hand, we took the sum from $n=1$ to $n=\rho_1$ in Eq. (76). This may not seem quite consistent with the assumption $n>>1$. Let us discuss the point more carefully and assume that in Eq. (76) we take the sum from

$$n=\frac{2}{10}\rho_1 \text{ to } n=\rho_1, \quad (79)$$

both values are much larger than unity, as needed. Then our sum becomes

$$\sum_{0.2\rho_1}^{\rho_1} n=\int_{0.2\rho_1}^{\rho_1} n dn=\frac{\rho_1^2}{2}\left(1-\frac{4}{100}\right), \quad (80)$$

which makes a very small correction. Combining conditions (78) and (79), we obtain

$$\theta_s \geqslant \frac{4}{n} \geqslant \frac{20}{\rho_1}, \quad \rho_1=ka=\frac{2\pi a}{\lambda}, \quad (81)$$

or

$$\theta_s=K\frac{\lambda}{a}, \quad K\approx 3,$$

and θ_s must still be much smaller than π. Here again let us be satisfied with an error 4/100,

$$\frac{20\lambda}{2\pi a} \leqslant \frac{4\pi}{100},$$

or approximately $\lambda/a \leqslant 0.04=1/25$, hence

$$\rho_1=\frac{2\pi a}{\lambda} \geqslant 160 \quad \frac{a}{\lambda}>25. \quad (82)$$

When conditions (81) and (82) are satisfied, the energy flow of the radiation scattered over angles comprised between θ_s and π corresponds within a few per cent per energy flow in the incident wave over a cross section πa^2. The secondary radiation inside an angle θ_s, if it were alone, would correspond to an energy flow measured by another equal cross section πa^2.

An attempt is made in Appendix III to rebuild the total field by the superposition of the direct and secondary waves, for these small angles. When

$$\theta<\theta_s, \quad \theta_s=\frac{2}{\rho_1}=\frac{\lambda}{\pi a}, \quad (83)$$

the total field is very small, but there is a bright fringe of interference along the z axis, when

$$\theta<\theta_F, \quad \theta_F=\frac{2}{\rho}=\frac{\lambda}{\pi R}, \quad (84)$$

where a is the radius of the sphere and R the distance of observation. This bright fringe along the axis is a well-known experimental fact, usually explained on the basis of Huygens' principle; it is obtained here on the basis of rigorous theory of electromagnetism.

The curve drawn on Fig. 5 represents an attempt at summarizing our results, but the calculations developed in the present report do not give enough information to enable us to draw a complete and precise graph.

No experimental device can measure the scattered radiation down to a deviation zero, hence there will always be a lower limit θ_s to the deviation angle. The assumption made in (81) and (82) corresponds to

$$\theta_s=\frac{4}{100}\pi \text{ or } \theta_s=7°.$$

The measurements are performed for all ratios λ/a and plotted in Fig. 5, with $\rho_1=2\pi(a/\lambda)$ as parameter and the scattering cross section Q_s as ordinate. For small values of ρ_1 (up to $\rho_1\approx 20$) the small finite value of θ_s does not play any appreciable role, and the scattering cross section Q_s is practically equal to the value computed by Stratton (taking $\theta_s=0$). It oscillates up and down, and finally reaches a value near $2\pi a^2$. For shorter wave-lengths (larger ρ_1 values), a transition will be observed, and the curve will go down to the limit πa^2 which corresponds to geometrical optics. The final value πa^2 is practically reached for $\rho_1\approx 160$ in our typical example.

We did not obtain any information about the shape of the curve in the intermediate region, but the curve should exhibit a few sharp maxima and minima, since our limit $\theta_s>K(\lambda/a)$ where $K=3$ includes all the first classical diffraction rings [Eq. (4)].

APPENDIX I

Some Formulas Connected with Legendre Functions

Legendre functions $P_n^1(x)$ satisfy the following equation[3]

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dP_n^1}{d\theta}\right)+\left(n(n+1)-\frac{1}{\sin^2\theta}\right)P_n^1=0. \quad (85)$$

This formula can be used in connection with the integral

$$I_2=\int_{\theta_s}^{\pi-\epsilon_2}\left(\frac{P_n^1 P_m^1}{\sin\theta}+\sin\theta\frac{dP_n^1}{d\theta}\frac{dP_m^1}{d\theta}\right)d\theta. \quad (86)$$

Integration by parts on the second term yields

$$I_2=\left|\sin\theta P_m^1\frac{dP_n^1}{d\theta}\right|_{\theta_s}^{\pi-\epsilon_2}+\int_{\theta_s}^{\pi-\epsilon_2}P_m^1\left[\frac{P_n^1}{\sin\theta}-\frac{d}{d\theta}\left(\sin\theta\frac{dP_n^1}{d\theta}\right)\right]d\theta$$

$$I_2=\left|\sin\theta P_m^1\frac{dP_n^1}{d\theta}\right|_{\theta_s}^{\pi-\epsilon_2}+n(n+1)\int_{\theta_s}^{\pi-\epsilon_2}P_n^1P_m^1\sin\theta d\theta. \quad (87)$$

If θ_s and ϵ_2 are zero, the conditions for orthonormalization of the P_n^1 give

$$I_2=\begin{cases}0 \text{ if } n\neq m\\ n(n+1)\dfrac{2}{2n+1}\dfrac{(n+1)!}{(n-1)!}=\dfrac{2n^2(n+1)^2}{2n+1}\end{cases} \quad (88)$$

if $n=m$.

The Legendre function P_n^1 can be represented by the expansion[25]

$$P_n^1=\frac{n(n+1)}{2}\sin\theta\times\left[1-\frac{(n-1)(n+2)}{4}(1-\cos\theta)\cdots\right]. \quad (89)$$

This shows that P_n^1 is zero for $\theta=0$ (and $\theta=\pi$). When $\theta\to0$

$$P_n^1=\frac{n(n+1)}{2}\theta\left[1-\frac{\theta^2}{6}-\frac{(n-1)(n+2)}{8}\theta^2\cdots\right]$$

$$\frac{dP_n^1}{d\theta}=\frac{n(n+1)}{2}\left[1-\frac{\theta^2}{2}-\frac{3}{8}(n-1)(n+2)\theta^2\cdots\right]. \quad (90)$$

$$R_n^1=-\frac{P_n^1}{\sin\theta}+\frac{dP_n^1}{d\theta} \quad (91)$$

is zero for $\theta=0$ and becomes appreciable (for large n) when $n^2\theta^2/4$ can no longer be neglected in comparison to 1 or when

$$\theta\sim\frac{2}{n}. \quad (92)$$

The Legendre function P_n^1 starts from zero at $\theta=0$ and exhibits a series of oscillations between $\theta=0$ and $\theta=\pi$. These oscillations can be represented by the approximate formula[26] (see Fig. 1).

$$P_n^1\approx-\left(\frac{2n}{\pi\sin\theta}\right)^{\frac{1}{2}}\sin\left[\left(n+\frac{1}{2}\right)\theta+\frac{3\pi}{4}\right], \quad (93)$$

when

$$n>>1, \quad \epsilon<\theta<\pi-\epsilon.$$

The limit ϵ corresponds to the neighborhood of the first zero of P_n^1. This first zero practically coincides with the second zero of the approximate expression (93), namely

$$(n+\tfrac{1}{2})\theta_1=\frac{5\pi}{4} \quad \theta_1\approx\frac{5\pi}{4n}, \quad (94)$$

a limit larger than (92) but of the same order of magnitude.

APPENDIX II

Some Elementary Results Concerning the Fourier Series and Conditions for Stationary Phase

The following Fourier series is known to yield a sawtoothed curve, with discontinuities at $x=0, \pm2\pi, \cdots$

$$f(x)=\sum_{p=1}^{\infty}\frac{\sin px}{p}=\begin{cases}\dfrac{-\pi-x}{2} & -2\pi<x<0\\ \dfrac{\pi-x}{2} & 0<x<2\pi.\end{cases} \quad (95)$$

The sudden jump from $-\pi/2$ to $+\pi/2$ when x crosses the zero point is of special importance for our discussion.

In the problem discussed in the report, the series to be computed is not exactly (95), but a similar series whose summation can be reduced to the case (95).

The series involved in Eq. (73) is

$$F(x)=\sum_{p=1}^{\infty}\cos^2p\psi\frac{\sin px}{p}=\frac{1}{2}\sum_{p=1}^{\infty}\frac{\sin px}{p}+\frac{1}{2}\sum_{p=1}^{\infty}\cos2p\psi\frac{\sin px}{p}=\frac{\partial\phi}{\partial n}, \quad (96)$$

where x is small and positive, while ψ is an angle between $-\pi/2$ and $-\pi/6$. The first series is given by (95) and the second series is split in two similar ones:

$$F(x)=\tfrac{1}{4}(\pi-x)+\frac{1}{4}\sum_{p=1}^{\infty}\left[\frac{\sin p(x+2\psi)}{p}+\frac{\sin p(x-2\psi)}{p}\right]. \quad (97)$$

Here, we must note that

$$x+2\psi<0$$
$$x-2\psi>0$$

[25] E. Jahnke and F. Emde, *Tables of Functions with Formulae and Curves* (B. G. Teubner, Leipzig and Berlin (G. E. Stechert & Company, New York), 1938), p. 111, Eq. (2). F. Emde, "Zur Berechnung der Reellen Nullstellen der Besselschen Zylinder funktione," Archiv. d. Math. u. Physik **24**, 239–250 (1916).

[26] See reference 25, p. 117, Eq. (6).

which leads to the two different cases of (95):

$$F(x)=\tfrac{1}{4}(\pi-x)+\tfrac{1}{8}[\pi-x-2\psi-\pi-x+2\psi]$$
$$=\frac{\pi}{4}-\frac{x}{2}, \quad (98)$$

a result that could have been guessed, since it practically amounts to taking $\cos^2 p\psi$ with its average value $\frac{1}{2}$ in the original formula (96).

APPENDIX III

Shadow Structure and Babinet Principle

It was noted in the general discussion in Section IX that a sharp distinction between incident wave and secondary wave is impossible to draw for small angles of deflection θ, and that the right way to attack the problem is to discuss the total field distribution as given by the superposition of incident and secondary waves.

Both waves are represented by the expansions given in Section VI, namely:

secondary wave (S), Eqs. (27);
plane incident wave (I), Eq. (27a). (99)

Both $h_n^1(\rho)$ and $j_n(\rho)$ functions were computed in Section VIII and their approximate expressions were used to obtain the a_n and b_n coefficients for a rather large metallic sphere ($a>\lambda$).

$$\rho_1=ka>>1, \quad \rho=kR, \quad k=2\pi/\lambda, \quad (100)$$

where a is the radius of the sphere and R the distance of observation. The discussion in Section VIII proved that the a_n, b_n coefficients become negligible when n is larger than ρ_1 [Case III, Eq. (65)], hence the expansion of the secondary wave will practically contain ρ_1 terms only.

We build up the resultant or total field by superimposing both expansions (99) of primary I and secondary S waves. The preceding remarks show that the ρ_1 first terms only will be affected by this superposition, while the remaining terms $n>\rho_1$ in the expansion of the plane primary wave remain unchanged.

$$\text{Total field}=\sum_{n=1}^{n=\rho_1}(I+S)+\sum_{n=\rho_1}^{\infty} I. \quad (101)$$

The last group of terms results directly from equation (27a) without any change. Let us now discuss the first group. We are especially interested in the behavior in the *shadow region*, which corresponds to small angles θ, and we assume condition (92) to be fulfilled,

$$\theta<\theta_s=2/\rho_1 \quad n\leqslant\rho_1. \quad (102)$$

Under such conditions we may use equation (90) and the following approximations are justified:

$$\frac{\partial P_n^1}{\partial\theta}\approx\frac{P_n^1}{\sin\theta}\approx\frac{1}{\theta}P_n^1=\frac{n(n+1)}{2}. \quad (103)$$

This simplifies the general expansions (27). If we study the field at a great distance, the radial component E_R is negligible and we are left with the θ and ϕ components only:

$$E_\theta=E_0e^{-i\omega t}\sum_{n=1}^{\rho_1} i^n\frac{2n+1}{n(n+1)}\frac{P_n^1}{\rho\theta}\times\begin{bmatrix}+a_n\rho h_n-ib_n[\rho h_n]'\\+\rho j_n-i[\rho j_n]'\end{bmatrix}\cos\phi,$$

$$E_\phi=-E_0e^{-i\omega t}\sum_{n=1}^{\rho_1} i^n\frac{2n+1}{n(n+1)}\frac{P_n^1}{\rho\theta}\times\begin{bmatrix}+a_n\rho h_n-ib_n[\rho h_n]'\\+\rho j_n-i[\rho j_n]'\end{bmatrix}\sin\phi. \quad (104)$$

This refers only to the first group of terms in (101). Terms in a_n, b_n, h_n represent the contribution of the secondary wave, while the j_n terms come from the direct wave. The brackets are identical in both formulas. Let us discuss their values for the problem stated in Section VIII, Case I.

$$n<\rho_1$$
$$\psi_n=\rho j_n=\cos\phi_n$$
$$a_n=-\cos\phi_{1n}e^{-i\phi_{1n}}$$
$$\phi_{1n}=\rho_1-\frac{\pi}{2}(n+1)$$
$$\rho_1=2\pi\frac{a}{\lambda}>>1$$
$$\zeta_n=\rho h_n=e^{i\phi_n}=-i[\rho h_n]'$$
$$b_n=-i\sin\phi_{1n}e^{-i\phi_{1n}}$$
$$\phi_n=\rho-\frac{\pi}{2}(n+1). \quad (105)$$

The primes indicate $\partial/\partial\rho$ derivatives, which amount to $\partial/\partial\phi_n$, and ϕ_n,ϕ_{1n} should not be confused with the angular coordinate ϕ.

$$a_n\zeta_n-ib_n\zeta_n'=-\cos\phi_{1n}e^{i(\phi_n-\phi_{1n})}-i\sin\phi_{1n}e^{i(\phi_n-\phi_{1n})}=-e^{i\phi_n}, \quad (106)$$

hence the bracket of formula (104) reduces to

$$-e^{i\phi_n}+\cos\phi_n+i\sin\phi_n=0.$$

This means that the first group of terms in (101) drops out entirely and that we obtain the following *very important result*,

$$\theta<\theta_s=\frac{2}{\rho_1}=\frac{\lambda}{\pi a} \quad \rho_1=\frac{2\pi a}{\lambda}. \quad (107)$$

$$\text{Total field}\approx\sum_{n=\rho_1}^{\infty}(\text{incident field terms}).$$

In the *shadow region*, the resulting field is obtained by *erasing* the *first* ρ_1 *terms* in the expansion of the incident wave.

$$E_\theta \approx \frac{E_0 e^{-i\omega t}}{\rho} \sum_{n=\rho_1}^{\infty} i^n \frac{2n+1}{n(n+1)} \times \left[\psi_n \frac{P_n^1}{\sin\theta} - i\psi_n' \frac{\partial P_n^1}{\partial \theta} \right] \cos\phi$$

$$E_\phi \approx -\frac{E_0 e^{-i\omega t}}{\rho} \sum_{n=\rho_1}^{\infty} i^n \frac{2n+1}{n(n+1)} \times \left[\psi_n \frac{\partial P_n^1}{\partial \theta} - i\psi_n' \frac{P_n^1}{\sin\theta} \right] \cos\phi, \quad (108)$$

where $\psi_n = \rho j_n$.

Here we may attempt to use again the Debye approximations (see Section VIII), and state

$$\begin{aligned} n << \rho \quad & \psi_n \approx \cos\phi_n = \cos\left[\rho - \frac{\pi}{2}(n+1)\right] \\ n >> \rho \quad & \psi_n \approx 0, \end{aligned} \quad (109)$$

hence sums in (108) extend practically from

$$n = \rho_1 \text{ to } n = \rho. \quad (110)$$

Equations (108) through (110) represent an approximate solution for the field in the shadow region, as defined by (107). The P_n's are oscillating functions, with an order of magnitude $(n/\sin\theta)^{\frac{1}{2}}$, according to (108), whenever θ does not become too small. Hence for such moderately small angles the sums (108) should remain very small and yield a very low light intensity.

A different situation arises when θ becomes *extremely small*

$$\theta < \theta_F = \frac{2}{\rho} = \frac{\lambda}{\pi R}$$

$$\theta_F = \frac{a}{R} \theta_a \quad (111)$$

where R is the distance of observation and a the radius of the sphere. We may now use (103) again in sum (108) and obtain

$$E_\theta \approx \frac{E_0 e^{-i\omega t}}{\rho} \sum_{n=\rho_1}^{\infty} i^n (n+\tfrac{1}{2})(\psi_n - i\psi_n'), \quad (112)$$

and a similar formula for E_ϕ. The upper limit in the sum is still of the order ρ, but because of the $(n+\frac{1}{2})$ factor, terms increase up to the limit, and it is impossible to use Debye's approximation, since the most important contribution to the sum is now given by terms $n \approx \rho$, for which no practical approximation was obtained (Section VIII, Case II). Let us try a more accurate method, starting from the rigorous formula,[27]

$$e^{i\rho} = \sum_{n=0}^{\infty} i^n (2n+1) j_n(\rho) \quad \gamma = 0. \quad (113)$$

Next we use the recurrence formula,[28]

$$j_n' = \frac{n}{2n+1} j_{n-1} - \frac{n+1}{2n+1} j_{n+1} \approx \frac{1}{2}(j_{n-1} - j_{n+1}) \quad n >> 1. \quad (114)$$

Turning back to (112),

$$\frac{1}{\rho}(\psi_n - i\psi_n') = j_n - \frac{i}{\rho}[\rho j_n]' \approx j_n - \frac{i}{2}(j_{n-1} - j_{n+1}) \quad \rho >> 1. \quad (115)$$

For large n and ρ values, the sum in (112) reduces to

$$\sum_{n=\rho_1}^{\infty} i^n n(\psi_n - i\psi_n') \approx \sum_{n=\rho_1}^{\infty} \{ i^n n j_n + \tfrac{1}{2}[i^{n-1}(n-1) j_{n-1} + i^{n+1}(n+1) j_{n+1}] \} \quad (116)$$

since

$$n + \tfrac{1}{2} \approx n \approx n-1 \approx n+1.$$

With the same approximation, (113) yields

$$\tfrac{1}{2} e^{i\rho} = \sum_{0}^{\infty} i^n n j_n. \quad (117)$$

In the three sums (116), small n values contribute very little. If and when we can add to the sums the small missing terms corresponding to $0 < n < \rho_1$, all three sums are identical with (117) and finally we obtain

$$\theta < \theta_F = \frac{\lambda}{\pi R},$$

$$E_\theta \approx E_0 e^{-i\omega t + i\rho}, \quad (118)$$

which yields a bright fringe all along the z axis.

This discussion leads to some interesting *conclusions* that shall be based upon our formulas (99), (101) and (108). Let us consider the expansions (104). They represent the field in a plane wave passing through a circular aperture of radius $a >> \lambda$. The series is limited to $\rho_1 = ka$ terms. It can be used only for small values of the angle of deflection θ, and this formula does not take into account the special properties of the iris. It represents a common approximation, for small angles, for all kinds of diaphragms. Large angle deflection will be entirely different for different diaphragms but small angle diffraction should be practically insensitive to the special properties of the iris.

[27] See reference 2, p. 409, Eq. (56).
[28] See reference 2, p. 406, Eq. (34).

If this point of view be adopted, then equations (107) and (108) represent the old "*Babinet's principle*"[7] that the field (in and near the shadow) is simply the incident field of the plane wave, minus the field of a plane wave passing through a hole of radius a. These two fields are mixed together in the shadow region, but they disentangle themselves at very large distances (Fig. 2) and the energy flux in the Babinet field corresponds to a cross section πa^2, a result that checks completely with the discussion of Section II.

Measurement of the Complex Conductivity of an Ionized Gas at Microwave Frequencies*

FRED P. ADLER**
California Institute of Technology, Pasadena, California
(Received May 13, 1949)

The positive column of a glow discharge is placed along the axis of a cylindrical cavity excited in the TM_{010} mode. The transmission of 3-cm waves through the cavity and the shift in resonant frequency are observed as a function of discharge current. It is shown that from these measurements values of the complex conductivity, $\sigma_r+i\sigma_i$, of the electron gas can be calculated. Curves of the measured conductivity components as functions of pressure and current are given. Using a theoretical formula for the conductivity (see reference 6) values of electron density can in turn be calculated from both σ_r and σ_i. Langmuir probe studies are carried out to check the results obtained, and adequate agreement is found.

I. INTRODUCTION

A KNOWLEDGE of the conductivity of an ionized gas is of interest in several phases of microwave work, such as the design of TR tubes,[1] as well as in many problems connected with the study of the ionosphere.[2] Some work has been done on discharges supported by microwave energy,[3] but published measurements of conductivity[4] are restricted to frequency ranges below 10^9 c.p.s.

The present paper describes a determination of conductivity at a wave-length of about 3 cm. A d.c. mercury vapor glow discharge was used to supply the conducting gas. This has the advantage that the electron temperature and density can be determined by probe studies, using Langmuir's theory;[5] in conjunction with the theoretical equation for conductivity[6] these measurements allow one to make a check of the values for the conductivity obtained experimentally.

II. EXPERIMENTAL APPARATUS

A schematic diagram of the apparatus is shown in Fig. 1. The cavity used (Fig. 2) was designed to resonate near 3 cm in the TM_{010} mode. Coupling to the wave guides was through circular irises; small openings at the centers of the top and bottom walls allowed a quartz discharge tube to extend axially through the cavity. Since in the mode used the cavity wall currents fall to zero at the centers of the end plates, these apertures have no appreciable disturbing effect. The quartz tube contained a mercury glow discharge whose vapor pressure was varied between 1 mm Hg and 18 mm Hg.

The high frequency excitation was obtained from a klystron tube frequency-modulated at a 1000-c.p.s. rate; the same modulation was also applied to the horizontal sweep of an oscilloscope. The output from the cavity was picked up by a crystal, amplified, and applied to the vertically deflecting plates of the oscilloscope. The resulting trace was thus a Q-curve, whose height was proportional to the transmission in power through the cavity since a square law crystal was used. A calibrated wave meter placed between cavity and crystal introduced small "pips" on the trace by means of which the

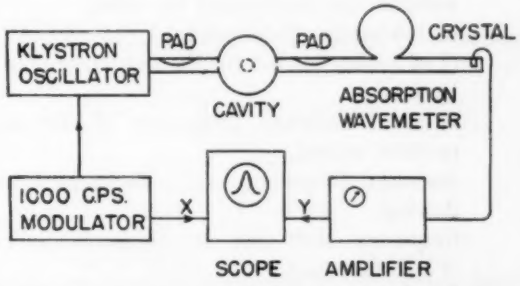


FIG. 1. Schematic diagram of experimental arrangement.

---

* Part of a thesis to be submitted by the author to the Graduate School at the California Institute of Technology in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

** General Electric Charles A. Coffin Fellow, 1947–1949.

[1] Margenau, McMillan, Dearnley, Pearsall, and Montgomery, Phys. Rev. **70**, 349 (1946).

[2] L. H. R. Mimno, Rev. Mod. Phys. **9**, 1 (1937).

[3] M. A. Herlin and S. C. Brown, Phys. Rev. **74**, 291 (1948); S. C. Brown and J. J. McCarthy, Rev. Sci. Inst. **19**, 851 (1948). For investigations of electrodeless discharges at lower than microwave frequencies, see: G. I. Babat, J. Inst. Elect. Eng., Part III, 27 (1947); L. Rhode, Ann. d. Physik **12**, 585 (1932).

[4] A. Szekely, Ann. d. Physik **3**, 112 (1929); E. V. Appleton and F. W. Chapman, Proc. Phys. Soc. (London) **44**, 246 (1932).

[5] I. Langmuir, Gen. Elec. Rev. **27**, 449, 538, 616 (1924).

[6] H. Margenau, Phys. Rev. **69**, 508 (1946). For an extension of this theory, taking into account inelastic electron collisions, see also: H. Margenau, Phys. Rev. **73**, 297 (1948); H. Margenau and L. M. Hartman, Phys. Rev. **73**, 309 (1948); T. Holstein, Phys. Rev. **70**, 367 (1946).

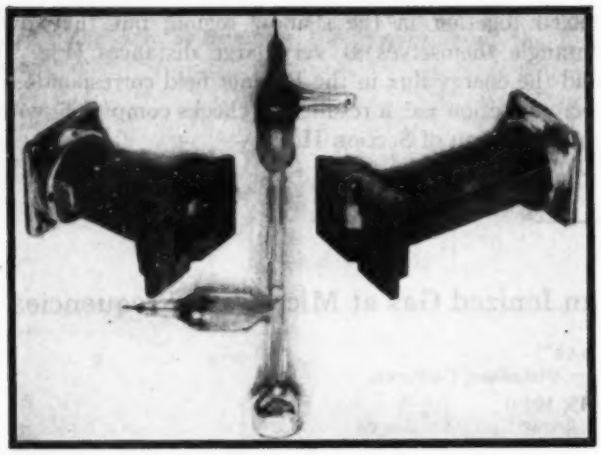


FIG. 2. Cavity and discharge tube

resonant frequency and the Q of the cavity could be measured. As shown below, measurement of these three quantities—effective Q, change in transmission, and change in resonant frequency with discharge current—suffices to determine both real and imaginary components of the conductivity.

For the purpose of carrying out probe studies a second discharge tube was built identical to the one used in the cavity, except for a fine probe wire extending radially into the middle portion.

III. THEORETICAL RELATIONS

A. Nomenclature

t . . . . . transmission in power through the cavity, i.e., ratio of output power from the cavity to available power.
t_r . . . . . relative transmission, i.e., ratio of transmission t with discharge current flowing to zero-current transmission.
a . . . . . radius of cavity (1.15 cm).
h . . . . . height of cavity (1.5 cm).
b . . . . . radius of conducting column (0.175 cm).
$\sigma_r+i\sigma_i$ . . complex conductivity of ionized mercury vapor, mhos/meter.
Q . . . . . ratio of resonant frequency to band width between half-power points; specifically the Q of the unloaded cavity.
Q_L . . . . . Q of cavity with loading due to coupling to the external circuit through the irises.
Q_i . . . . . an iris loading Q defined by $1/Q_L=1/Q+1/Q_i$.
Q' . . . . . Q of cavity loaded by the discharge (but not by irises).
ω_0 . . . . . nominal resonant frequency of the cavity radians/second.
ω_r . . . . . resonant frequency with discharge current flowing.
Δf . . . . . frequency shift due to discharge in c.p.s. $=1/2\pi(\omega_r-\omega_0)$.
p . . . . . vapor pressure, mm Hg.
T . . . . . electron temperature, °K.
n . . . . . electron density, electrons/cm^3.
x_1 . . . . . a dimensionless parameter $=(m\omega^2\lambda^2/2kT)$, λ being the electronic mean free path; for mercury at 8460 Mc: $x_1=1.2\times10^6p^{-2}T^{-1}$.
ϵ_1 . . . . . $\epsilon_0(1+\sigma_i/\omega\epsilon_0)$; cf. Eq. (8).
k_1 . . . . . $\omega_r(\mu\epsilon_1)^{\frac{1}{2}}$.
k_2 . . . . . $\omega_r(\mu\epsilon_0)^{\frac{1}{2}}$.
k . . . . . $2.405/a$.

B. Calculation of the Real Component of Conductivity

The transmission through the cavity is given by[7]

$$t=(1-Q_L/Q)^2\doteq(1+Q_i/Q)^{-2}. \tag{1}$$

For a cavity loaded by wall losses only, Q may be expressed in terms of an equivalent shunt resistance R along the cavity axis as[8]

$$Q=0.00531\frac{a}{h}R. \tag{2}$$

The small additional loading due to the presence of the quartz tube can be assumed to be lumped into R, giving a slightly lower effective Q. The conducting column is now assumed to introduce an additional shunt resistance R' placed in parallel with R:

$$R'=\frac{h}{\pi b^2}\cdot\frac{1}{\sigma_r}. \tag{3}$$

Thus the resultant Q becomes now

$$Q'=0.00531\frac{a}{h}\cdot\frac{RR'}{R+R'} \tag{4}$$

or using (2) to substitute for R:

$$Q'=\frac{Q}{1+188Qh/R'a}. \tag{5}$$

This then is the expression which must replace the Q appearing in (1) whenever the cavity is loaded by the discharge; carrying through this substitution, and using (3) to write R' in terms of σ_r, the resultant expression may be manipulated into the following factored form:

$$t=(1+Q_i/Q)^{-2}\left(1+592\frac{b^2}{a}Q_L\sigma_r\right)^{-2}. \tag{6}$$

The first bracket is recognized as the transmission with no discharge current flowing; using this as a base the second bracket may be defined as a relative transmis-

[7] G. L. Ragan, *Microwave Transmission Circuits* (McGraw-Hill Book Company, Inc., New York, 1948), p. 654.
[8] J. G. Brainerd *et al.*, *U.H.F. Techniques* (McGraw-Hill Book Company, Inc., New York, 1942), p. 331.

sion, t_r, in terms of which σ_r can be expressed explicitly as:

$$\sigma_r = 0.00169 \frac{a}{b^2} \frac{1}{Q_L} (t_r^{-1} - 1) \text{ mhos/meter.} \quad (7)$$

t_r is readily determined, being simply the ratio of the height of the Q-curve trace with discharge current flowing to that with the glow turned off; for convenience and accuracy a meter was actually used which could be calibrated to read t_r directly. Since all quantities on the right-hand side of (7) can be measured, σ_r may be found.

C. Calculation of the Imaginary Component of Conductivity

From Maxwell's Field equation

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + (\sigma_r + i\sigma_i)\mathbf{E} = \sigma_r \mathbf{E} + i\omega\epsilon_0 \left(1 + \frac{\sigma_i}{\omega\epsilon_0}\right)\mathbf{E},$$

the out-of-phase conductivity component σ_i is seen to result in an effective dielectric constant

$$\epsilon_1 = \epsilon_0 \left(1 + \frac{\sigma_i}{\omega\epsilon_0}\right). \quad (8)$$

(Note: Rationalized m.k.s. units are used.) Since σ_i can be shown to be negative, being due to the inertial lagging of the electrons behind the applied field, the dielectric constant is reduced, decreasing the equivalent capacity of the cavity; thus the effect of the glow discharge is to raise the resonant frequency, as shown by Eq. (12) below.

To treat the cavity as a boundary value problem, it is divided into two regions, where for $0 < \rho < b$, $\epsilon_1 = \epsilon_0(1 + \sigma_i/\omega\epsilon_0)$, while for $b < \rho < a$ the dielectric constant is still ϵ_0. No account is taken of the narrow region filled by the quartz tube which introduces a small constant frequency shift. For the inner region the expressions for the fields will be of the same form as for the usual TM_{010} mode:

$$E_{z1} = J_0(k_1\rho), \quad H_{\phi 1} = (\omega\epsilon_1/k_1) J_1(k_1\rho). \quad (9)$$

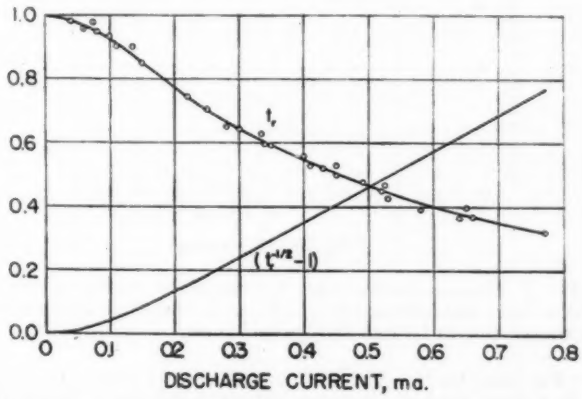


FIG. 3. Transmission characteristic at 1 mm Hg pressure.

In the outer region Bessel functions of the second kind need also be included, so that for $b < \rho < a$:

$$E_{z2} = AJ_0(k_2\rho) + BY_0(k_2\rho),$$
$$H_{\phi 2} = (\omega\epsilon_0/k_2)[AJ_1(k_2\rho) + BY_1(k_2\rho)]. \quad (10)$$

Three boundary conditions apply: at $\rho = a$, $E_{z2} = 0$; at $\rho = b$, $E_{z2} = E_{z1}$ and $H_{\phi 1} = H_{\phi 2}$. Using two of these to eliminate the arbitrary constants A and B, the third one may be used to obtain the following relation which contains the desired resonant frequency implicitly in k_1 and k_2:

$$\frac{J_1(k_2b)Y_0(k_2a) - J_0(k_2a)Y_1(k_2b)}{J_0(k_2b)Y_0(k_2a) - J_0(k_2a)Y_0(k_2b)} = \left(\frac{\epsilon_1}{\epsilon_0}\right)^{\frac{1}{2}} \frac{J_1(k_1b)}{J_0(k_1b)}. \quad (11)$$

Both sides of this equation may be linearized over a small region near $\epsilon_1 = \epsilon_0$, $k_1 = k_2$, $\omega_r = \omega_0$; after some numerical work this leads to the following first-order approximation for the particular case at hand:

$$\sigma_i = -1.33 \times 10^{-9} \Delta f \text{ mhos/meter.} \quad (12)$$

An alternate method of deriving a relation between frequency shift and σ_i makes use of the variation principle for electromagnetic problems (analogous to Hamilton's principle in mechanics), *viz.*, that

$$\int\int_{\text{time volume}} (\tfrac{1}{2}\epsilon E^2 - \tfrac{1}{2}\mu H^2) dv dt$$

have a stationary value of zero. Since all time variations are sinusoidal, and H may be replaced from Faraday's

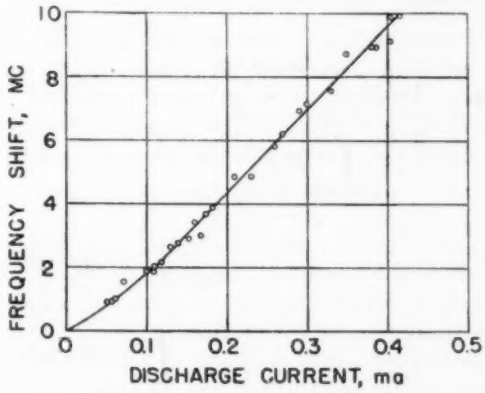


FIG. 4. Frequency shift at 2 mm Hg pressure.

Law, $\nabla \times \mathbf{E} = -i\omega\mu\mathbf{H}$, this relation can be rewritten as

$$\omega^2 = \frac{\int_{\text{vol}} \frac{1}{\mu} (\nabla \times \mathbf{E})^2 dv}{\int_{\text{vol}} \epsilon E^2 dv}. \quad (13)$$

Assuming that the perturbation of the electric field due

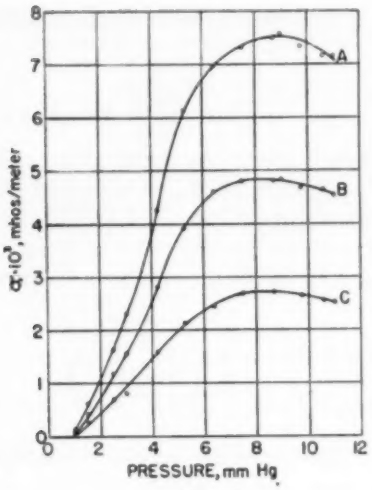


FIG. 5.

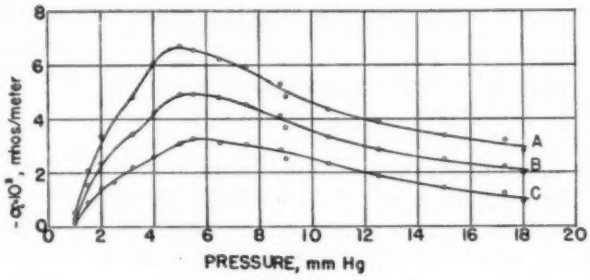


FIG. 6.

FIGS. 5 AND 6. Conductivity *versus* pressure: A. discharge current = 130 μa; B. discharge current = 100 μa; C. discharge current = 70 μa.

to the discharge is not appreciable, it is seen that the effect on the resonant frequency of any small changes in the dielectric constant can be expressed by

$$\frac{\omega_r}{\omega_0}=\left[\frac{\int_{\text{vol}}\epsilon_0 E^2\cdot dv}{\int_{\text{vol}}\epsilon E^2\cdot dv}\right]^{\frac{1}{2}}, \tag{14}$$

where in this case ϵ is given by Eq. (8) over the region $0<\rho<b$. Since the change in ϵ due to σ_i is of the order of only 0.5 percent, one may write $1/(1+\sigma_i/\omega\epsilon_0)^{\frac{1}{2}}=1-\sigma_i/2\omega\epsilon_0$, obtaining:

$$\frac{\omega_r-\omega_0}{\omega_0}=-\frac{1}{2}\frac{\int_{\text{vol}}(\sigma_i/\omega\epsilon_0)E^2dv}{\int_{\text{vol}}E^2dv}$$

$$=\frac{-\sigma_i}{2\omega\epsilon_0}\frac{\int_0^b J_0^2(k\rho)\rho d\rho}{\int_0^a J_0^2(k\rho)\rho d\rho}. \tag{15}$$

Hence

$$\Delta f=-\frac{1}{2\pi}\frac{\sigma_i}{2\omega_0\epsilon_0}\left(\frac{b}{a}\right)^2\frac{J_0^2(kb)+J_1^2(kb)}{J_1^2(ka)}. \tag{16}$$

Making the proper numerical substitutions, Eq. (16) is found to agree within one percent with Eq. (12).

D. Theoretical Expression for Conductivity

A theoretical expression for the complex conductivity of an ionized gas has been derived by H. Margenau.[6] For mercury vapor at 8460 Mc, the effective resonant frequency at the cavity with the tube inserted, this expression becomes:[9]

$$\sigma_r+i\sigma_i=4.38\times10^{-10}\frac{n}{pT^{\frac{1}{2}}}K_2(x_1)$$

$$-i4.81\times10^{-7}\frac{n}{p^2T}K_{3/2}(x_1), \tag{17}$$

where

$$K_n(x_1)\equiv\int_0^\infty\frac{x^ne^{-x}}{x+x_1}dx;$$

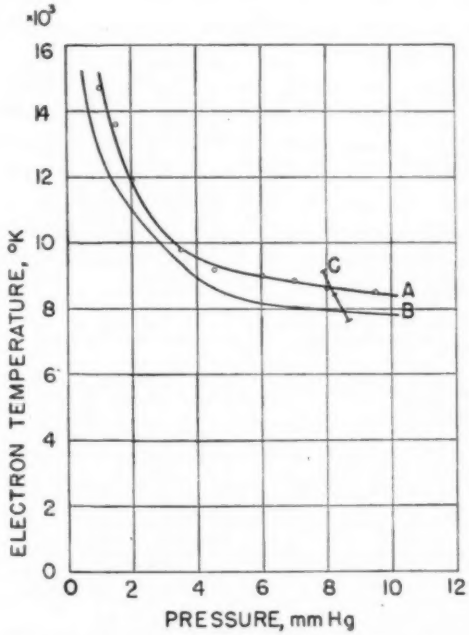


FIG. 7. Electron temperature: A. from probe studies; B. theoretical curve (see reference 11); C. from microwave measurements (corresponding to 7.9 mm Hg $\leq p_0 \leq$ 8.7 mm).

[9] The value for the mean free path of mercury was taken from J. D. Cobine, *Gaseous Conductors* (McGraw-Hill Book Company, Inc., New York, 1941), p. 23.

K_2 and $K_{3/2}$ have been plotted by Margenau. The contribution of the mercury ions can be shown to be negligible due to their large mass; hence only effects due to the electrons are actually observed.

The real component of the conductivity falls to zero at both very high and very low pressures; a maximum is predicted[6] for

$$\omega\lambda = 6.8\times 10^{13}(kT)^{\frac{1}{2}} \text{ (c.g.s.)} \qquad (18)$$

which corresponds very nearly to the condition of the mean free time of an electron between collisions being equal to the period of the electromagnetic wave. For mercury at the frequency used (16) becomes:[9]

$$p_0^2 T = 570{,}000, \qquad (19)$$

where p_0 is the pressure in mm Hg at which σ_r maximizes.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Typical curves of t_r, $(t_r^{-1}-1)$, and Δf as a function of discharge current are shown in Figs. 3 and 4. Since σ_r and σ_i have been found to be proportional to $(t_r^{-1}-1)$ and Δf, respectively, these curves show that there exists a very nearly linear relationship between conductivity and current. This same conclusion is arrived at by the analytical conductivity expression, Eq. (17), which shows σ to be proportional to the electron density n, which in turn varies approximately linearly with discharge current. The slightly curved portion in the low current region is most probably due to a similar variation at n with current.[10]

By taking transmission and frequency shift data at various pressures, plots of conductivity *vs.* pressure with current as parameter may be obtained through use of Eqs. (7) and (12) (Figs. 5 and 6). The real conductivity component is seen to maximize at pressures around 8 mm Hg. Corresponding values of electron temperature may be calculated from this by Eq. (19); these are shown in curve *C*, Fig. 7. Also shown there are values of electron temperature found by probe measurements as well as a theoretical temperature plot.[11]

Figure 6 shows a maximum occurring also in the imaginary conductivity component. Now if σ_i is plotted against pressure at constant electron density, rather than constant current, Eq. (17) predicts a monotonic decrease with pressure starting from a maximum value of $e^2n/m\omega$. Hence the observed falling off at lower pressures is apparently due to a similar decrease of electron density at these pressures. To check this assumption values of σ_r/n and σ_i/n were calculated from Eq. (17), using the measured temperature shown in Fig. 7B. By dividing these two quantities into the measured values of σ_r and σ_i, respectively, the electron density is obtained. Curves of density calculated in this manner are plotted in Fig. 8; a systematic discrepancy of about 20 percent is noticed between values found from using σ_r and the ones obtained from σ_i. As a further check electron density was also found from probe measurements (see Fig. 8). This method involved plotting probe current squared against probe voltage; at sufficiently high values of probe current this results in a straight line from whose slope electron density may be determined.[12] At the large probe currents required the discharge became unfortunately quite unstable at pressures above 2 mm Hg, so that a complete check of the electron density curves could not be carried through.

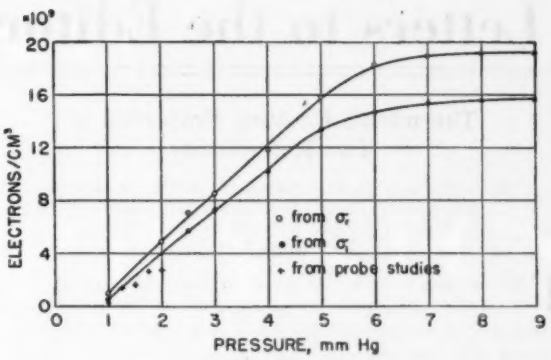


FIG. 8. Electron density *versus* pressure.

The agreement found in comparisons of experiment with theory, as shown in Figs. 7 and 8, is probably as good as may be expected from a study of this kind.

V. ACKNOWLEDGMENT

The author is indebted to Dr. M. D. Fiske of the General Electric Research Laboratory who first suggested a similar study in 1946, and to Dr. J. D. Cobine, of the same laboratory, under whose guidance this project was initiated. The bulk of the work done at the California Institute of Technology under the able supervision of Professors W. H. Pickering and S. S. Mackeown.

[10] Curvatures similar to the ones observed can be seen also in plots of n *vs.* current in the following: A. Szekely, Ann. d. Physik **3**, 112 (1929), Fig. 8; I. Langmuir, see reference 6, Table XIII, v. Engel and Steenbeck, *Elektrische Gasentladungen* (Verlag Julius Springer, Berlin, 1932), Vol. I, p. 186, Vol. II, p. 110 (1934).

[11] v. Engel and Steenbeck, see reference 10, Vol. II, p. 85.

[12] I. Langmuir, see reference 6, Eq. (46).

Letters to the Editor

Thermionic Emitting Properties of Two Refractories

R. E. Haddad, D. L. Goldwater, and F. H. Morgan
Cathode Research Group, Bartol Research Foundation of the Franklin Institute, Swarthmore, Pennsylvania
September 8, 1949

THE authors have investigated the thermionic emitting properties of two more refractory carbides.[1] These are tantalum carbide and titanium carbide. Both were applied as coatings on tungsten wire, and tested in guard-ringed diodes.

The titanium carbide coating is formed by painting a suspension of the powder in butyl alcohol onto the filament. In the case of tantalum carbide, the powdered material is suspended in nitrocellulose binder. Thin coatings of fairly uniform diameter are obtained.

Values of A and b are obtained for the Richardson equation $I_s = AT^2 e^{-b/T}$, ($b = 11{,}605\phi$). The values obtained are: for tantalum carbide, $A = 2.(1)$ amp./deg. K^2 cm^2, $\phi = 3.0(5)v$. For titanium carbide, $A = 2.(5)$ amp./deg. K^2 cm^2, $\phi = 3.3(5)v$.

The emission available does not seem to be of practical interest for either of these materials. At 1727°C$_b$, maximum space-charge limited emission from titanium carbide is about 0.4 amp./cm^2. At 1672°C$_b$, maximum space-charge limited emission from tantalum carbide is about 0.2 amp./cm^2.

The spectral emissivities of both materials were obtained, as necessary to the experiment. For titanium carbide, this emissivity, as measured at 0.655μ, is 0.96±0.04. Brightness temperatures for this material may hence be regarded as true centigrade temperatures. For tantalum carbide, the emissivity determined is 0.75±0.03.

[1] Zirconium carbide having been previously reported; Haddad, Goldwater, and Morgan, J. App. Phys. **20**, 886 (1949).

Papers by Hadrian D. Stahl

G. H. Metson
Post Office Engineering Department, Research Branch, London, England
August 29, 1949

I READ with surprise the two papers by Stahl published recently.[1] The layout and content of the papers are almost identical with one published by Stahl and Wagener.[2] I appreciate that this German paper, published during the war years, might be unknown in the States. The object of my note, therefore, is to draw attention to Wagener's share of the work which is probably at least equal to that of Stahl.

[1] H. A. Stahl, J. App. Phys. **20**, 1, 8 (1949).
[2] H. A. Stahl and S. Wagener, Zeits. f. Tech. Physik **24**, 280 (1943).